Etude structurale des protéines du complexe réplicatif du virus de la rougeole

David Karlin

UNIVERSITE DE LA MEDITERRANEE
AIX-MARSEILLE II
FACULTE DES SCIENCES DE LUMINY

N° attribué par la bibliothèque

THESE
pour obtenir le grade de
DOCTEUR DE L'UNIVERSITE DE LA MEDITERRANEE
Ecole Doctorale des Sciences de la Vie et de la Santé
Discipline : Sciences de la Vie
présentée et soutenue publiquement
par
David KARLIN
le 27 Mai 2002

Etude structurale des protéines du complexe réplicatif du virus de la rougeole

Directeur de thèse : Dr Bruno CANARD

JURY

Mr. Pr Jean Louis ROMETTE (Président du jury)
Mr. Dr Denis GERLIER (Rapporteur)
Mr. Pr Joseph CURRAN (Rapporteur)
Mr. Pr Laurent ROUX (Examinateur)
Mr. Dr Bruno CANARD (Directeur de thèse)
M^{me} Dr Sonia LONGHI (Co-directeur de thèse)

Table des Matières

À mes parents.

RESUME

Le virus de la rougeole est un virus à ARN négatif, membre de la famille des *Paramyxoviridae*. Le complexe de réplication viral comprend trois protéines principales : la polymérase à ARN ARN-dépendante (L), la nucléoprotéine (N) et la phosphoprotéine (P). Cette thèse présente une étude structurale, par des méthodes biochimiques et biophysiques, de la phosphoprotéine et de la nucléoprotéine produites dans la bactérie *Escherichia coli*. J'ai étudié les déterminants de la polymérisation de N et identifié un variant intéressant pour l'étude de N par cristallographie aux rayons X. J'ai aussi mis au point la purification d'un complexe de N et P, prélude à son étude structurale. Par ailleurs, j'ai montré que la partie N-terminale de P est intrinsèquement désordonnée. J'ai pu étendre ce résultat aux protéines P de nombreux virus apparentés par une étude bio-informatique. J'ai ensuite démontré l'importance du désordre structural au sein du complexe réplicatif des *Paramyxoviridae* (à la fois par son abondance et d'un point de vue fonctionnel). Au-delà de leur intérêt pratique immédiat, ces résultats indiquent que la machinerie réplicative de ces virus pourrait constituer un système modèle pour l'étude du désordre structural dans les protéines. De plus, ils soulèvent de nombreuses questions et sont donc un prélude à une refonte de notre vision du complexe réplicatif de ces virus dont l'étude revêt une importance majeure en santé humaine.

ABREVIATIONS

Sigle	*Signification de l'abréviation*	*Signification en anglais de l'abréviation (s'il y a lieu)*
aa	Acide aminé	
ARN	Acide ribonucléique	
ARNm	Acide ribonucléique messager	
C	Protéine C	
CDV	Virus de la maladie de Carré	Canine distemper virus
F	Protéine de fusion	
H	Protéine d'attachement	
L	Grande sous-unité de la polymérase, produit du gène L	Large subunit of the polymerase
M	Protéine de matrice	
OMS	Organisation mondiale de la santé	
N	Nucléoprotéine	
N_{CORE}	Partie N-terminale de la nucléoprotéine	
N_{TAIL}	Partie C-terminale de la nucléoprotéine	
N°	Forme soluble de la nucléoprotéine	
N^{NUC}	Forme assemblée de la nucléoprotéine	
P	Phosphoprotéine	
PCT	Partie C-terminale de la phosphoprotéine	
PMD	Domaine de multimérisation de P	P multimerization domain
PNT	Partie N-terminale de la phosphoprotéine	
PONDR	Logiciel de prédiction des régions intrinsèquement désordonnées	Predictor of naturally disordered regions
PX	Domaine de liaison de P à la nucléocapside	Protein X
RMN	Résonance magnétique et nucléaire	
RNP	Complexe ribonucléoprotéique	
RV	Virus de la rage	Rabies virus
RT-PCR	Réaction de polymérisation en chaîne couplée à une réaction de transcription inverse.	Reverse-transcription-polymerase chain reaction
SAXS	Diffusion des rayons X aux petits angles	Small-angle X-ray scattering
SSPE	Pan-encéphalite sclérosante sub-aigüe	Subacute sclerosing panencephalitis
V	Protéine V	

HISTORIQUE ET REPERES

Classification du virus de la rougeole

Le virus de la rougeole appartient au genre des *Morbillivirus*, dans la famille des *Paramyxoviridae*, au sein de l'ordre des *Mononegavirales*. Ce dernier comprend de nombreux virus pathogènes pour l'homme, tels que le virus des oreillons, le virus de la rage, et le virus d'Ebola, ainsi que des virus émergents comme les virus Nipah ou Hendra (voir tableau 1).

Les *Paramyxoviridae* n'infectent que des vertébrés ; en particulier, les *Morbillivirus* n'infectent que des mammifères, aussi bien terrestres que marins (tableau 1). À l'opposé, les *Rhabdoviridae* infectent un large spectre d'hôtes, des insectes aux plantes (tableau 1).

Malgré leur importance en termes de santé humaine et animale, nos connaissances sur les mécanismes moléculaires de la transcription et de la réplication des *Mononegavirales* sont très limitées. Elles ont pour la plupart été obtenues à partir d'études conduites sur des virus modèles, comme le virus de Sendai (SeV) ou le virus de la stomatite vésiculaire (VSV). Ceux-ci n'ont pourtant aucun impact sur la santé humaine, puisqu'ils infectent respectivement les souris et divers vertébrés supérieurs.

Les éléments des chapitres suivants sur l'historique, la pathogenèse et l'épidémiologie de la rougeole sont tirés principalement de la revue "Measles Virus" de Diane Griffin (Griffin, 2001) et du livre "Plague's Progress" d'Arno Karlen (Karlen, 1995), tous deux remarquables. Le lecteur intéressé s'y réfèrera

.

Tableau 1

ORDRE	FAMILLE	SOUS-FAMILLE	GENRE	ESPÈCE	Abréviation	HOTE(S)
Mononegavirales	*Paramyxoviridae*	*Paramyxovirinae*	*Morbillivirus*	**Measles virus (virus de la rougeole)**	MV	Mammifères Terrestres et marins : Homme
				Rinderpest virus	RPV	Bovidés
				Canine distemper virus (virus de la maladie de Carré)	CDV	Chien, lion, furet
				Phocine distemper virus	PDV	Phoque
				Dolphin Morbillivirus	DMV	Dauphin
				Porpoise Morbillivirus	PMV	Marsouin
				Peste-des-petits-ruminants virus	PPRV	Petits ruminants
			Respirovirus	Sendai virus (SeV)	Sev	Souris
				Human parainfluenza virus type 1	hPIV-1	Homme
				Bovine parainfluenza virus type 3	bPIV-3	Boeuf
				Human parainfluenza virus type 3	hPIV-3	Homme
			Rubulavirus	**Human parainfluenza virus type 2**	hPIV-2	Homme
				Mumps virus (virus des oreillons)	MuV	Homme
				Simian virus type 5	SV5	Singe
				Simian virus type 41	SV41	Singe
				Tioman virus	TiV	Chauve-souris
			Henipavirus	**Nipah virus** (1)	NiV	Chauve-souris, porc, homme
				Hendra virus (1)	HeV	Chauve-souris, cheval, homme
			Avian genus	Newcastle disease virus	NDV	Oiseaux (pigeon …)
			Non classés	Salem virus	SalV	Cheval
				Tupaia paramyxovirus	TPMV	Tupaia (musaraigne arboricole)
		Pneumovirinae	*Meta-pneumovirus*	**Human metapneumovirus**	n.r.	Homme
				Turkey rhinotracheitis virus	TRTV	Dinde, poulet
			Pneumovirus	**Human respiratory syncytial virus**	hRSV	Homme
				Bovine respiratory syncytial virus	bRSV	Boeuf

ORDRE	FAMILLE	GENRE	ESPÈCE	Abréviation	HOTES
Mononegavirales	*Rhabdoviridae*	*Lyssavirus*	**Rabies virus (virus de la rage)**	RV	Homme, chien, renard
		Vesiculovirus	Vesicular stomatitis Indiana virus	VSV	Vertébrés supérieurs (rongeurs)
		Ephemerovirus	Bovine ephemeral fever virus	BEFV	Boeuf
		Novirhabdovirus	Infectious haematopoetic necrosis virus	IHNV	Truite, saumon
		Cytorhabdovirus	Lettuce necrotic yellows virus	LNYV	Plantes
		Nucleorhabdovirus	Sonchus yellow net virus	SYNV	Plantes
		Non classifiés	Sigma virus (2)	n.r.	Drosophile
	Filoviridae	"Ebola-like viruses"	**Zaire Ebola virus** (1)	n.r.	Homme, singe, rongeurs
		"Marburg-like viruses"	**Marburg virus** (1)	n.r.	Homme, singe
	Bornaviridae	*Bornavirus*	**Borna disease virus**	BV	Homme, cheval

Tableau 1 (légende)

Les virus pathogènes pour l'homme sont signalés en gras. Si la traduction en français du nom d'un virus ne découle pas immédiatement du nom anglais, le nom français courant est mis entre parenthèses. Le virus prototype de chaque genre est indiqué en premier. La liste des hôtes n'est pas exhaustive. L'hôte principal ou d'intérêt principal en santé humaine ou vétérinaire est mentionné en premier. n.r. : non renseigné.

Notes :

(1) : virus émergents

(2) Sigma virus: virus des drosophiles célèbre pour induire chez ces dernières une mystérieuse sensibilité au CO_2 *transmissible* qui a joué un grand rôle dans l'étude des arbovirus (voir le remarquable livre sur l'histoire de la virologie de Claude Chastel : (Chastel, 1992)).

Historique

Le virus de la rougeole est sans doute dérivé du virus qui cause la maladie de Carré chez les chiens, (CDV, voir Tableau 1), que ceux-ci avaient eux-mêmes hérité de leurs ancêtres, les loups. En effet, le virus de la rougeole induit une immunité qui persiste toute la vie, et il n'a aucun réservoir animal. Il nécessite donc à tout moment l'existence d'un nombre minimal de personnes réceptives (au moins 7 000). Cet apport ne peut provenir que de naissances ou de migrations. Ainsi, au cours des derniers siècles, la rougeole s'est éteinte dans des îles ayant moins de 500 000 habitants, à moins d'être réintroduite de l'extérieur. Ce chiffre illustre bien son extrême contagiosité. Pour ces raisons, on pense que la rougeole n'a pu apparaître qu'avec les premières grandes civilisations urbaines, il y a 3 000 ou 4 000 ans, à la fin de l'âge de bronze, ou au début de l'âge du fer.

Comme d'autres maladies contagieuses, l'apparition de la rougeole dans des populations vierges pouvait prendre une apparence spectaculaire, avec des symptômes et une mortalité terrifiantes. Nous n'en avons que peu de témoignages, mais les dernières épidémies dans des populations vierges en donnent une idée. Quand la rougeole frappa les îles Fidji, vers 1870, elle y provoqua de 20 à 30% de morts parmi les populations natives. La rougeole eut une virulence comparable quand elle apparut en Afrique de l'Ouest, à Hawaii, et aux Samoa.

Les écrits médicaux anciens ne font pas la distinction entre les différentes infections exanthématiques (c'est-à-dire caractérisées par des éruptions cutanées), comme la rougeole, la variole, la scarlatine, la rubéole, la varicelle ou l'érysipèle. Le médecin arabe Abu Bir, connu sous le nom de Rhazes de Bagdad, a été le premier à distinguer la rougeole de la variole au IX^{ème} siècle, même s'il pensait qu'elles étaient deux variantes d'une même maladie. De fait, la rougeole voyageait souvent de pair avec la variole, une autre maladie très contagieuse et dévastatrice. Des épidémies répétées de maladies exanthématiques ont eu lieu en Europe et en Orient depuis le début de notre ère jusqu'en l'an 1200, date à laquelle la rougeole est mentionnée pour la première fois comme une maladie infantile. En particulier, la rougeole serait arrivée en France avec l'invasion des sarrasins au VIII^{ème} siècle. Au XVI^{ème} siècle, la rougeole était devenue une maladie familière, avec une mortalité relativement faible. Dans la littérature de l'époque, la rougeole et les autres exanthèmes étaient appelés "morbilli" d'après l'italien "petites maladies", pour les distinguer de la peste, *il morbo*. C'est alors que les conquistadors la propagèrent dans le Nouveau Monde, où elle devait tuer des dizaines de millions d'indigènes pendant 200 ans, contribuant à l'effondrement de plusieurs civilisations.

Pathogenèse

La rougeole est principalement une maladie infantile humaine. Une période d'incubation de 10 à 14 jours est suivie par un épisode de 2 à 3 jours de fièvre, coryza, et de conjonctivite, suivie par l'apparition d'une éruption cutanée caractéristique et bien connue. Le début de cette éruption coïncide avec l'apparition de la réponse immunitaire et le début de l'élimination du virus par l'organisme. La guérison s'accompagne en général d'une immunité à vie.

Le virus, transmis par aérosol, entre par les voies respiratoires. Depuis la trachée, le virus progresse jusqu'aux nœuds lymphatiques locaux, probablement par l'intermédiaire de cellules dendritiques ou de macrophages pulmonaires. De là il se répand par les voies sanguines dans plusieurs organes, véhiculé principalement par les monocytes. On observe rarement du virus libre dans le sang, même si on peut détecter de l'ARN viral par RT-PCR.

La rougeole induit une immunosuppression transitoire, qui fut découverte parce qu'elle diminuait le test bien connu d'hypersensibilité à la tuberculine, qui se pratique pour vérifier l'immunité d'un individu contre la tuberculose. Du fait de cette immunosuppression, la rougeole peut provoquer la rémission de maladies auto-immunes telles que l'arthrite rhumatoïde juvénile, mais elle cause aussi une susceptibilité accrue vis-à-vis d'infections latentes comme la tuberculose.

Une complication létale de la rougeole a fait l'objet de beaucoup d'études au niveau moléculaire, en dépit de sa faible importance clinique (un cas sur un million) : la panencéphalite sclérosante sub-aigüe (SSPE). Les modifications des protéines virales observées dans la SSPE ont permis d'éclairer le rôle de certaines de ces protéines (Gombart *et coll.*, 1992; Hirano *et coll.*, 1993; Hirano *et coll.*, 1992; Reutter *et coll.*, 2001; Suryanarayana *et coll.*, 1994).

Epidémiologie

La rougeole est une des maladies les plus contagieuses. On a estimé qu'un seul individu peut mettre en contact avec le virus de 20 à 25 personnes, et que dans une population vierge, les trois-quarts développeront la rougeole ! La compréhension des bases épidémiologiques de l'infection doit beaucoup à des études d'épidémies insulaires, et en particulier à celle de Peter Panum, un jeune médecin danois envoyé aux îles Faroe en 1846 pendant une grande épidémie. Celui-ci comprit la nature hautement contagieuse de la maladie, la période d'incubation de 14 jours, l'immunité à vie présente chez les habitants plus âgés, et postula une transmission par voies respiratoires.

La couverture vaccinale a profondément changé l'épidémiologie de la rougeole. Dans les pays où la couverture est très forte, la plupart des cas de rougeole proviennent d'un rejet du vaccin pour des raisons culturelles ou religieuses. Dans le monde entier, la rougeole cause entre 750 000 et 1 000 000 de morts par an, contre environ 30 en France (où il y a 60 000 cas par an, soit environ 1 mort sur 2000 cas).

L'approche vaccinale

En théorie, la rougeole est un virus idéal pour une politique d'éradication par vaccination. Il n'y a qu'un seul sérotype, la plupart des cas sont identifiables cliniquement, il n'y a pas de réservoir animal, et un vaccin est disponible. L'éradication de la rougeole aux Etats-Unis (ou tout au moins de sa transmission endémique) a été programmée en 1979, mais ce but n'a été atteint que récemment. En 1989, l'OMS a désigné un premier objectif à l'échelle mondiale, le contrôle du virus, avant son éradication (Omer, 1999; Ruff, 1999).
C'est la mise en culture du virus (souche Edmonston) en 1954 qui permis le développement de vaccins efficaces. À l'origine, il y avait deux types de vaccin : les vaccins reposant sur des virus inactivés et les virus atténués dits "vivants". Les vaccins reposant sur des virus inactivés (par la formaline ou l'éther-tween) ne conféraient qu'une protection temporaire, et furent la cause de cas atypiques graves de rougeole. En revanche, la stratégie des virus atténués a permis le développement de vaccins très efficaces. La célèbre souche vaccinale Edmonston B est fabriquée par un passage de la souche Edmonston sauvage sur des cellules primaires rénales et amniotiques humaines, suivie par son atténuation par adaptation à des œufs embryonnés de poulets puis à des fibroblastes d'embryons de poulets. Des variantes de ce protocole ont conduit à d'autres souches vaccinales plus efficaces et mieux tolérées. Citons notamment la souche Schwarz (qui tient lieu de standard dans beaucoup de pays) et la souche Moraten utilisée aux Etats-Unis. L'âge à partir duquel on pratique le vaccin varie suivant le risque dans le pays considéré. Dans les pays à risque, l'immunisation se fait vers 9 mois, dès que les anticorps maternels anti-rougeole ont disparu. Cependant, la séroconversion suite à une vaccination précoce peut être moins bonne, aussi ailleurs l'immunisation a lieu vers 12-15 mois. On estime que l'efficacité moyenne d'une dose unique de vaccin est de 80-95%; une seconde dose améliore cette efficacité. C'est la stratégie en deux doses qui aurait permis l'élimination de la rougeole endogène en Finlande et aux Etats-Unis.

Les stratégies actuelles pour éradiquer la rougeole reposent sur la distribution du vaccin à de grands nombres en très peu de temps grâce à des campagnes de masse. Des programmes globaux ont augmenté la couverture vaccinale dans le monde, et en conséquence, la mortalité a notablement diminué. Cependant, cette stratégie rencontre un certain nombre d'obstacles, comme le rejet du vaccin pour des raisons culturelles ou religieuses. De plus, la diffusion du vaccin reste difficile dans de nombreux endroits, notamment à cause de guerres civiles (par exemple au Rwanda en 1993-94) mais aussi de réticences psychologiques ou de problème de conservation du vaccin. Le vaccin lyophilisé est relativement stable, mais une fois reconstitué, il s'inactive rapidement à température ambiante. Aussi faut-il maintenir une chaîne du froid (difficile quand on reste 3 semaines en brousse sans électricité). De plus, on ne dispose toujours pas de vaccin adapté aux très jeunes enfants. Enfin, le rôle des

infections asymptomatiques et des individus immunisés d'une manière incomplète dans la pérennité du virus de la rougeole n'est pas clair.

Traitement

Il n'y a pas de traitement antiviral standard pour la rougeole. L'administration de hautes doses de vitamine A durant un cas aigu de rougeole diminue la mortalité, et son utilisation est recommandée par l'OMS dans les pays sévèrement touchés par la rougeole.

L'utilisation de **médicaments antiviraux** (non disponibles actuellement) serait un complément très utile à cette approche vaccinale en permettant de juguler rapidement une épidémie naissante. Or le système de réplication viral constitue la cible principale des agents antiviraux développés à ce jour. Citons notamment quelques antiviraux classiques dirigés contre la machinerie réplicative - les virus cibles sont indiqués entre parenthèses - : la stavudine (virus de l'immunodéficience humaine (VIH)) ; la lamivudine (virus de l'hépatite B) ; l'acyclovir (virus de l'herpès). Un préalable à la conception rationnelle de tels agents antiviraux est la compréhension du mécanisme de réplication du virus de la rougeole, qui fait l'objet de notre travail.

STRUCTURE DU VIRION ET ORGANISATION DU GENOME

STRUCTURE DU VIRION

Les *Paramyxoviridae* sont des virus enveloppés généralement sphériques, avec un diamètre compris entre 150 et 350 nm. Leur enveloppe est constituée d'une bicouche lipidique qui provient de la membrane plasmique de la cellule hôte, et dans laquelle sont insérées des glycoprotéines d'origine virale : la protéine d'attachement (H) et la protéine de fusion (F). H doit son nom au fait qu'elle peut provoquer l'agglutination d'érythrocytes (Hématies) de primates. L'enveloppe repose sur un squelette protéique constitué par la protéine de matrice (M). Chaque virion de la rougeole contient plusieurs complexes ribonucléoprotéiques (RNP) formés chacun par une molécule d'ARN viral génomique ou antigénomique associée aux protéines N, P, et L (Lund *et coll.*, 1984) (voir chapitre suivant). Une représentation schématique d'un *Paramyxoviridae* est présentée figure 1A, en regard d'une image en microscopie électronique figure 1B.

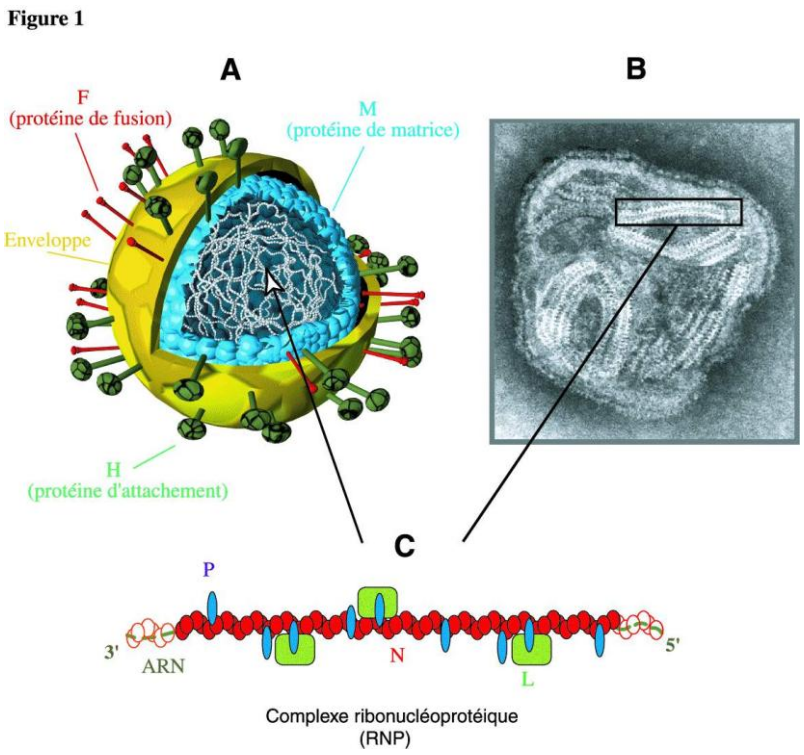

Figure 1. Structure du virion

A. Vue d'artiste d'un *Paramyxoviridae*.

Le virion est représenté comme une sphère idéale. En réalité, il est pléiomorphique (voir figure 1B). Le dessin n'est pas à l'échelle. Le virus est formé d'une matrice sur laquelle repose une bicouche lipidique d'enveloppe. Deux glycoprotéines virales, F et H, sont insérées dans cette enveloppe. A l'intérieur du virion, les RNP viraux (voir figure 1C), sont représentés à travers la coupe de l'enveloppe. Cette vue d'artiste est reproduite avec l'aimable permission de Jaspal Randhawa (Pneumovirus Laboratory, Dept. of Biological Sciences, University of Warwick, Coventry, Royaume-Uni).

B. Particule virale vue en microscopie électronique.

On distingue l'enveloppe externe, et à l'intérieur de la particule, des nucléocapsides à l'aspect caractéristique dit "en arêtes de hareng" (L et P ne sont pas visibles en microscopie électronique). (Image : Linda Stannard, Department of Medical Microbiology, University of Cape Town, Afrique du Sud).

C. Complexe ribonucléoprotéique (RNP) viral

Les composants du RNP sont dépeints comme : ovales rouges, nucléoprotéine (N); rectangle vert, polymérase (L); ovales bleus, multimères de phosphoprotéine (P); ligne pointillée verte, ARN génomique. Celui-ci est encapsidé par N dans une nucléocapside hélicoïdale, représentée symboliquement ici. Les extrémités 3' et 5' de l'ARN encapsidé sont représentés par transparence à travers les nucléoprotéines aux deux extrémités de la nucléocapside, pour plus de clarté. Noter qu'il n'y a pas d'extrémité libre.

Des études sur SeV et hPIV-1, des Respirovirus, ont montré qu'il existe des multimères de P libres, non complexés avec L. Ainsi, dans des RNP engagés dans la transcription, P est regroupé en amas autour de L, mais on trouve aussi des multimères de P distants des molécules de L, ce que nous avons représenté ici.

ORGANISATION DU GENOME

L'organisation du génome est très conservée chez les *Paramyxoviridae*, et nous détaillerons seulement celle du virus de la rougeole. Le génome viral est constitué de 15894 nucléotides. Il existe deux types d'ARN viraux de cette longueur, l'un de polarité négative (génome), et l'autre de polarité positive (antigénome) (voir figure 2). La région 3' extracistronique du génome, contenant 55 nucléotides, est appelée région "leader". Une région semblable, appelée "trailer" ou "leader (-)" se trouve à l'extrémité 5' du génome. Le génome du virus de la rougeole contient 6 gènes: N, P, M, F, H et L (figure 3). Le gène P dirige la synthèse de 4 protéines (voir chapitre suivant). La polymérase ne peut pénétrer dans la matrice génomique (-) et antigénomique (+) qu'au niveau de l'extrémité 3', via une interaction avec le promoteur viral, qui est toujours actif (Lamb and Kolakofsky, 2001). Chaque gène comprend une région régulatrice de départ de la transcription (qui est elle-même transcrite en ARN) et une région d'arrêt et de polyadénylation du transcrit, se terminant par une courte répétition d'uridines (pour revue : (Galinski, 1991; Horikami and Moyer, 1995)). Les différents gènes sont séparés par une région intergénique (de 3 nucléotides pour le virus de la rougeole) (figure 3).

La longueur du génome de la plupart des *Paramyxovirinae* est un multiple de 6 ; seuls des virus respectant cette règle (dite "règle de 6") peuvent se propager (Calain and Roux, 1993). Ceci est dû au fait que chaque nucléoprotéine interagit avec précisément 6 nucléotides de l'ARN viral, et au sein de N les sites de liaison de chacun des 6 nucléotides ne sont pas équivalents (on parle de "phase hexamérique") (Kolakofsky *et coll.*, 1998; Vulliemoz and Roux, 2001). Ainsi, les promoteurs viraux de l'ARN de SeV ou de SV5 contiennent des éléments placés dans une phase hexamérique bien précise (Murphy *et coll.*, 1998; Murphy and Parks, 1999; Tapparel *et coll.*, 1998).

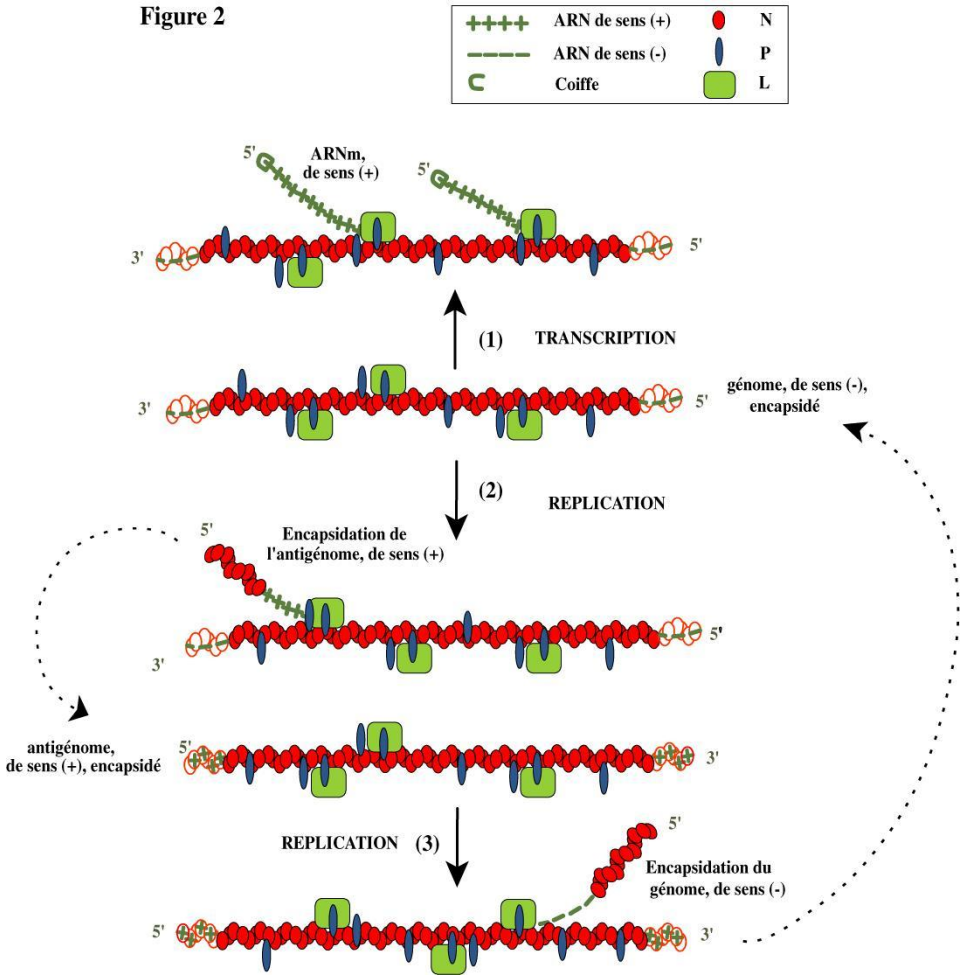

Figure 2. Cycle réplicatif.

Les conventions sont les mêmes qu'en figure 1C. L'ARN génomique est défini comme étant de polarité (-), et l'ARNm viral de polarité (+). Les ARN (+) sont représentés par des signes (+) verts, et les ARN (-) par des signes (-).

1) Transcription de l'ARN génomique en ARNm. Noter que les ARNm sont coiffés, un fait symbolisé par un demi-cercle à l'extrémité 5' de l'ARNm.

2) Réplication de l'ARN génomique en une copie complémentaire intégrale, l'antigénome, de polarité (+). L'antigénome est encapsidé en parallèle avec sa synthèse ; on ne sait pas si l'ARN naissant est encapsidé dès qu'il est synthétisé ou bien avec un léger retard, comme nous l'avons représenté ici pour plus de clarté.

3) L'ARN antigénomique est à son tour répliqué en un ARN génomique, qui est encapsidé en parallèle. Cet ARN génomique sert de matrice à un nouveau cycle de synthèse d'ARNm, appelé transcription secondaire (non montré).

Figure 3

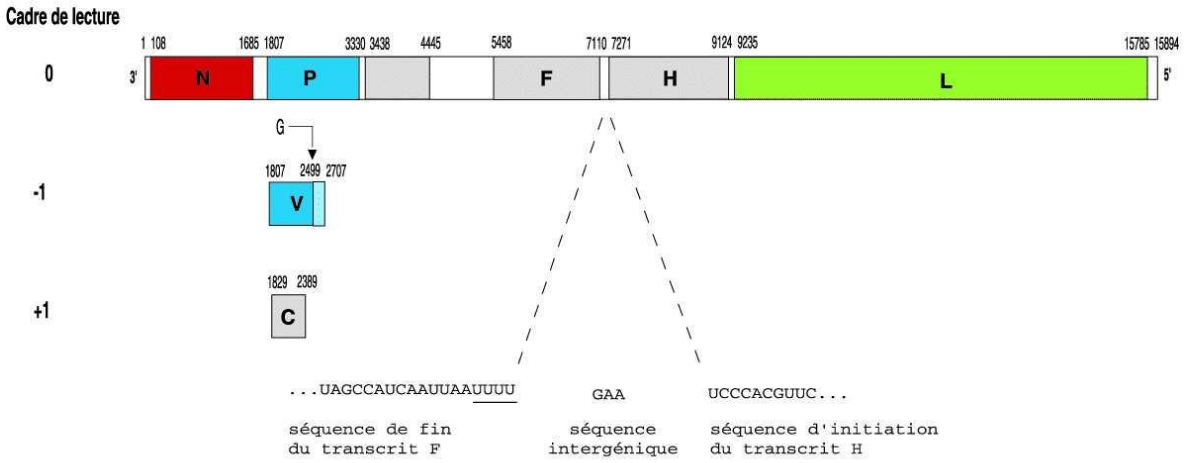

Figure 3. Organisation du génome du virus de la rougeole.

Haut : Organisation codante. Les 3 cadres de lecture (pris par référence au cadre de lecture de N) sont représentés. Les régions codant pour les protéines virales dans ce dernier cadre de lecture sont représentées en couleurs pour N, P, et L (rouge, bleu, et vert, respectivement) et en gris pour M, F, H. L'ajout co-transcriptionnel par la polymérase virale d'un nucléotide G en position 2499 crée un nouvel ARNm qui code pour la protéine V; celle-ci est composée d'une partie commune avec P (en bleu) et d'une région unique, issue du cadre de lecture -1 (en bleu clair). Dans le cadre de lecture +1, une initiation alternative de la transcription conduit à la synthèse de la protéine C (en gris).

Bas : Signaux régulateurs de la transcription.

Le signal d'arrêt de la transcription du gène F, suivi par la séquence intergénique entre F et H, puis le signal de départ de la transcription du gène H, sont indiqués. La courte séquence répétée d'uridines est soulignée.

La diversité d'expression du gène P

Chez les *Paramyxovirinae*, le gène P représente un exemple extraordinaire de compactage de l'information génétique. Il dirige la synthèse d'au moins deux protéines (et jusqu'à 8 dans le cas de SeV !) grâce à l'utilisation de cadres ouverts de lecture partiellement chevauchants. Ce décalage du cadre de lecture est produit par plusieurs mécanismes : initiation alternative, édition du transcrit, ou décalage ribosomal (pour revue: (Latorre *et coll.*, 1998)). En particulier, 4 protéines sont exprimées à partir du gène P chez tous les *Morbillivirus*. L'insertion co-transcriptionnelle d'un G supplémentaire en position 753 de l'ARNm de P crée un ARNm distinct dont le produit est la protéine V (figure 4). Celle-ci partage avec P toute la région traduite en amont de cette insertion, le domaine PNT (aa 1-230), mais possède un domaine unique riche en cystéines, hautement conservé en séquence, impliqué dans la liaison de deux atomes de zinc (figure 4). Une autre protéine, C, est produite par un mécanisme d'initiation alternative : le ribosome initie la traduction de l'ARNm de P sur un

codon situé en aval du codon initiateur principal et décalé dans le cadre de lecture (voir figure 3). Enfin, une autre protéine, R, est produite par un mécanisme de décalage ribosomal, avec une fréquence de seulement 2% (Liston and Briedis, 1995). Aucune étude n'a été réalisée sur R depuis son identification, et son importance fonctionnelle est inconnue. C et V sont deux protéines auxiliaires qui modulent la transcription et la réplication virales. Bien que leur fonction ne soit pas élucidée, elles semblent aussi jouer un rôle *in vivo* dans la pathogénicité, la virulence et l'apoptose (Escoffier *et coll.*, 1999; Patterson *et coll.*, 2000; Valsamakis *et coll.*, 1998).

Figure 4

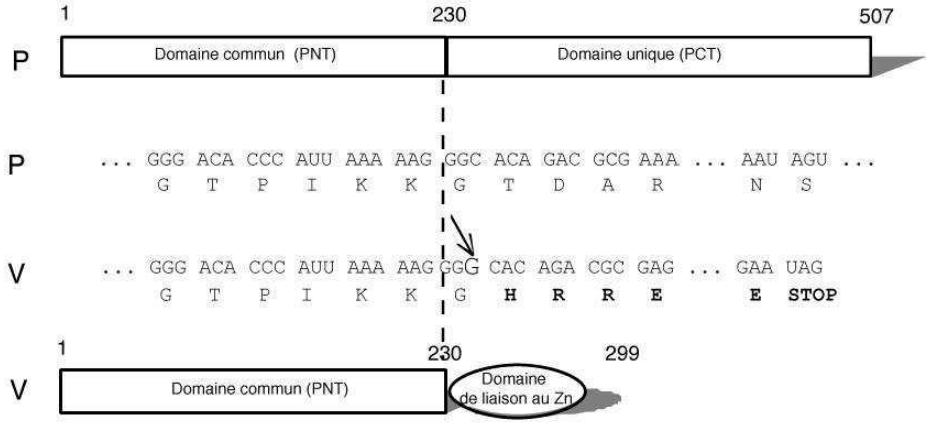

Figure 4. Production du transcrit de V à partir du gène P.

L'insertion (par "bégaiement" de la polymérase) d'un G en position 753 de l'ARNm de P (indiqué par une flèche) cause un décalage du cadre de lecture qui donne naissance à un nouveau transcrit, V. Les acides aminés spécifiés dans le nouveau cadre de lecture et distincts de ceux spécifiés par le cadre de lecture de V sont indiqués en gras. Dans le cadre de lecture de V, un nouveau codon STOP situé en amont du codon STOP de P interrompt la traduction; la protéine V possède ainsi 299 acides aminés. Les protéines V et P partagent un module N-terminal (PNT); leur partie C-terminale diffère - respectivement PCT pour P et Zn-BD (zinc-binding domain, domaine de liaison au zinc) pour V.

LE COMPLEXE RIBONUCLEOPROTEIQUE (RNP)

Le génome des *Paramyxoviridae* est constitué par de l'ARN simple brin, non segmenté et de polarité négative. L'ARN génomique doit donc être transcrit en ARNm complémentaire, et c'est ce dernier qui sert de matrice à la traduction des protéines virales (voir figure 2). L'ARN génomique est encapsidé par la nucléoprotéine (N) de manière à former une structure hélicoïdale appelée nucléocapside, qui sert de matrice à la transcription et à la réplication virale. Celles-ci sont effectuées par l'ARN-polymérase virale ARN-dépendante (L), en association avec la phosphoprotéine (P). Ce complexe ribonucléoprotéique (RNP) formé par l'ARN viral, N, P, et L est représenté en figure 1C. Dans le cas de SeV, chaque RNP est composé d'environ 2600 N, 300 P, et 50 L (Lamb and Choppin, 1977b).

La nucléocapside des *Paramyxoviridae* possède une structure hélicoïdale en forme de tube creux, présentant une morphologie particulière en "arête de hareng", bien visible sur la figure 1B. Cette hélice gauche mesure environ 1 µm de long pour un diamètre de 18 nm, avec un vide central de 4 à 5 nm de diamètre. Contrairement à la nucléocapside du virus de la mosaïque du tabac (TMV) à laquelle elle ressemble en microscopie électronique, la nucléocapside des *Paramyxoviridae* est relativement souple, ce qui lui permet d'être empaquetée dans le virion (qui mesure environ 150 nm de diamètre). L'hélice des *Paramyxoviridae* tels que le virus simien de type 5 (SV5) ou SeV peut adopter différents pas qui correspondent à des conformations plus ou moins compactes (Egelman *et coll.*, 1989; Heggeness *et coll.*, 1980; Heggeness *et coll.*, 1981) (figure 5). Dans la conformation la plus étendue de la nucléocapside de SeV (pas hélical de 37.5 nm contre 5.3 pour la forme la plus compacte), l'ARN viral est certainement exposé. Il est tentant d'imaginer qu'un tel phénomène de "désenroulement" pourrait intervenir dans la synthèse d'ARN pour permettre l'accès de la polymérase à l'ARN.

Figure 5

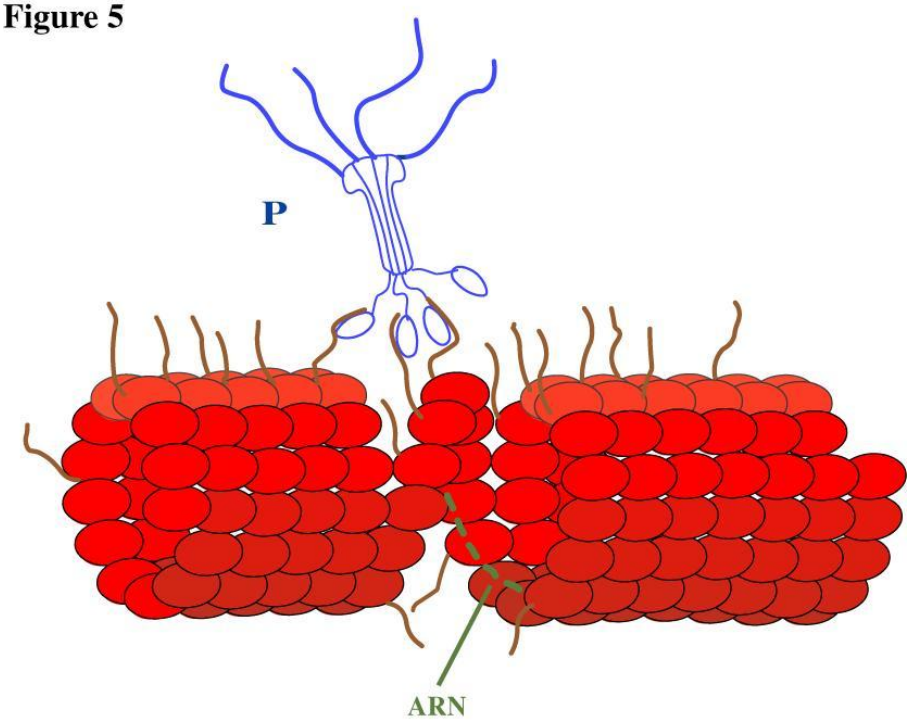

Figure 5. Changement de conformation hélicoïdale de la nucléocapside.
Représentation schématique d'une transition hélicoïdale de la nucléocapside du virus de la rougeole (voir texte). Dans cette représentation, elle comprend 13 nucléoprotéines par tour d'hélice ; il s'agit d'une hélice gauche. En réalité, il existe deux types de nucléocapsides, comprenant exactement 13.07 ou 13.41 monomères/tour (D. Bhella et P Yeo, communication personnelle). La partie N-terminale de N (le "cœur", N_{CORE}) est représentée par un ovale rouge, et la partie C-terminale de N (la "queue", N_{TAIL}) par une ligne brune. Seules quelques régions N_{TAIL} ont été représentées pour plus de clarté. Au centre de l'image, 2 tours de nucléocapside sont représentés dans une conformation complètement détendue ou "désenroulée", pour plus de clarté. En fait, en microscopie électronique plusieurs transitions hélicoïdales sont observées (4 dans le cas de SeV), avec des pas compris entre 5.3 et 37.5 nm (Egelman *et coll.*, 1989) ; il est possible que seul un relâchement minime de la nucléocapside (par exemple, pour un pas de 6.8 nm) suffise à exposer l'ARN.
Merci à Dave Bhella et à Paul Yeo pour leur image de reconstruction en microscopie électronique, qui a servi de support au dessin de la nucléocapside.

La nucléoprotéine, bien plus qu'une protéine structurale

Le génome de tous les virus à ARN négatif spécifie une nucléoprotéine qui se lie à l'ARN simple brin, N. Sa fonction première est d'encapsider le génome viral. Cependant, N est bien plus qu'une protéine structurale de liaison à l'ARN, mais fonctionne aussi comme une molécule adaptatrice entre le virus et les processus de la cellule hôte. Durant la transcription et la réplication, N interagit avec l'ARN, avec elle-même, et avec P seule ou complexée avec L. Lors de l'assemblage, elle interagit avec la protéine de matrice. Enfin, elle interagit aussi avec des facteurs cellulaires dont probablement certains composants du cytosquelette.

La phosphoprotéine, un rôle de pivot dans la transcription et la réplication

La phosphoprotéine joue un rôle central dans la transcription et la réplication. Tout comme N, elle remplit de nombreuses fonctions. Elle se lie à la nucléocapside, une propriété indispensable à la synthèse d'ARN à deux titres : d'une part, en tant que cofacteur de la polymérase (L), elle permet la liaison de cette dernière à la nucléocapside ; et d'autre part, sa liaison à N indépendamment de L joue un rôle non éclairci à ce jour.
De plus, P empêche l'auto-assemblage prématuré de N en formant avec elle un complexe (N°-P) qui sert de substrat à l'encapsidation. On pense que P joue un rôle actif dans ce processus, en permettant à la polymérase d'assembler précisément N sur l'ARN néosynthétisé.

Les rôles et les caractères structuraux de N et P sont présentés en détail dans la revue "Structural flexibility of the replicative complex of *Paramyxovirinae*".

Une enzyme-clé : la protéine L, une polymérase à ARN ARN-dépendante

Chez tous les virus à ARN négatif, l'ARN est encapsidé par la nucléoprotéine qui ne s'en dissocie sans doute jamais. Lors de la synthèse d'ARN, L doit donc interagir non pas simplement avec l'ARN matrice, mais avec un complexe (N:ARN). De plus, L possède un cofacteur indispensable à son activité : P. L présente donc des différences notables avec les polymérases à ARN ARN-dépendantes d'autres virus pour lesquels l'ARN nu sert de substrat, comme le virus de l'hépatite C. De fait, L possède les motifs communs à toutes les polymérases à ARN ARN-dépendantes, appelés A, B', C et D (O'Reilly and Kao, 1998), mais aussi de nombreuses autres régions sans homologie avec les polymérases à ARN ARN-dépendantes d'autres virus que les *Mononegavirales* (Poch *et coll.*, 1990). Sa grande taille (250 kDa) très supérieure à celle de ces dernières illustre bien ce fait.
Étant donné qu'aucune ARN-polymérase recombinante des *Paramyxoviridae* n'a été purifiée et étudiée *in vitro*, la cartographie des activités putatives de L provient de preuves indirectes. Les activités enzymatiques "classiques" attendues pour L (en complexe avec P) sont : coiffe des ARN (phosphatase spécifique, guanylyltransférase, méthyltransférase), activité de polyadénylation, d'édition du transcrit, adénosine 5'-triphosphatase (ATPase), kinase et/ou phosphatase, activité de type hélicase). Notons que la dernière activité pourrait être assurée par N, à l'image de la nucléoprotéine du virus de la grippe (Baudin *et coll.*, 1994), ce qui concorderait bien avec l'absence de motif canonique d'hélicase dans L. 6 motifs conservés (numérotés de I à VI) ont été trouvés dans L (Poch *et coll.*, 1990). Cependant, Ferron *et coll.* ont identifié un domaine methyltransférase impliqué dans la coiffe des ARN dans la partie C-terminale de L, precisant ainsi que le site de fixation de nucléotides (VI) identifié par Poch *et*

coll. dans ce dernier article était en fait un site de fixation de S-adenosyl-L-méthionine, le donneur de méthyl dans la réaction de coiffe (Ferron *et coll.*, 2002).

Chez les *Mononegavirales*, L doit interagir avec de nombreux autres partenaires que n'ont pas les polymérases à ARN ARN-dépendantes d'autres virus. L doit notamment :
- s'associer à l'autre sous-unité du complexe catalytique, P
- interagir avec le complexe (N:ARN) lors de la synthèse d'ARN, même si la liaison stable de L à la nucléocapside est assurée par P.
- reconnaître les séquences des promoteurs génomiques et antigénomiques pour initier la synthèse d'ARN.
- probablement interagir avec des molécules de P supplémentaires, présentes sur la nucléocapside en plus de leur association avec L (Curran, 1998).
- utiliser le complexe N°-P comme substrat d'encapsidation, et sans doute provoquer le relargage de N° par P lors de l'encapsidation (voir revue). Il est tentant d'imaginer que la **disponibilité** des complexes N°-P pour servir de substrat à l'encapsidation puis le relargage de N° par P puissent être régulés par des modifications post-traductionnelles de N ou de P (telles que la phosphorylation) effectuées par L. On a ainsi attribué une activité kinase à L de SeV, dont N et P seraient substrats, un résultat controversé (Einberger *et coll.*, 1990).

Il devrait être possible de cartographier certaines des régions responsables de ces fonctions en comparant les protéines L à travers différents genres viraux : ainsi, chez les *Rubulavirus* uniquement, L possède un site de liaison (transitoire ?) à N, bien qu'elle s'y associe aussi par l'intermédiaire de sa sous-unité P (Nishio *et coll.*, 2000). La région correspondante de L devrait être conservée chez les *Rubulavirus* mais pas dans les autres genres viraux.

Les seules données sur le rôle de P dans l'activité enzymatique de la polymérase proviennent d'une étude par protéolyse limitée de RNP de SeV, qui a montré que le domaine N-terminal de P, PNT, n'était pas nécessaire à la coiffe ou à la polyadénylation des ARN. Cependant, l'étude n'a pu déterminer si le domaine C-terminal de P, PCT, était impliqué dans ces activités (bien qu'on sache qu'il est indispensable à la stabilité de L) (Chinchar and Portner, 1981). Par ailleurs, des données non publiées de J. Curran (communication personnelle) suggèrent que N participe activement à la régulation de la terminaison de la transcription (voir chapitre "transition entre transcription et réplication").

Il ne fait aucun doute que notre connaissance de L peut encore nettement progresser grâce à l'emploi de la bio-informatique, mais des études structurales directes sont nécessaires pour élucider l'organisation de L. A l'heure actuelle, la seule protéine L qui soit purifiée en quantités suffisantes pour une étude cristallographique est celle de VSV, un *Rhabdoviridae*

(Das *et coll.*, 1998). Nous avons tenté au laboratoire (y compris dans le cadre de cette thèse) de produire L en baculovirus, ou dans la levure *P. Pastoris*, sans succès ; l'expression dans *E. coli* de sous-domaines putatifs de L définis par des études bio-informatiques n'a pour l'instant pas permis d'obtenir L en quantités souhaitées. D'autres approches qui sortent du cadre de cette thèse sont en cours au laboratoire.

Durant le cours de ma thèse, je me suis consacré à l'expression, à la purification et à la caractérisation des protéines virales (et/ou de leurs domaines) impliquées dans la transcription et la réplication, à savoir la **Nucléoprotéine (N)**, la **Phosphoprotéine (P)**, la **protéine auxiliaire V**, et la **polymérase (L)**. Les travaux présentés ici portent sur N et P.

LE CYCLE VIRAL

ENTREE DANS LA CELLULE ET DESASSEMBLAGE DU VIRUS

La première étape de toute infection virale est l'attachement du virus à des récepteurs cellulaires de surface, suivie par la pénétration du virus au travers de la membrane cellulaire (pour revue : (Young, 2001)). Le virus de la rougeole ne peut infecter que des cellules dérivées de primates supérieurs, et cette observation a permis l'identification d'un récepteur cellulaire du virus, CD 46, qui est un co-facteur du complément (Naniche *et coll.*, 1993). Des observations plus récentes indiquent que ce récepteur ne serait utilisé que par des souches vaccinales, le récepteur utilisé lors de l'infection par des virus sauvages étant en fait une molécule impliquée dans l'activation des lymphocytes, SLAM (signaling lymphocytic activation molecule), appelée aussi CDw150 (Ono *et coll.*, 2001; Tatsuo *et coll.*, 2001). Tous les *Morbillivirus* semblent faire usage de ce récepteur, auquel se lie H (Tatsuo *et coll.*, 2001). En accord avec son rôle dans l'activation des lymphocytes, la liaison de H à SLAM pourrait jouer un rôle dans l'immunosuppression partielle induite par la rougeole, une hypothèse controversée (Erlenhoefer *et coll.*, 2001; Hsu *et coll.*, 2001; Tanaka *et coll.*, 2002). Après attachement, la protéine de fusion (F) des *Paramyxoviridae* dirige la fusion entre l'enveloppe virale et la membrane cellulaire de la cellule-cible, un événement qui se déroule à pH neutre. F possède d'importants traits communs avec les protéines de fusion de virus très éloignés. Ainsi, F forme des homotrimères qui doivent être clivés par une protéase pour être actifs. Ce clivage découvre à l'extrémité N-terminale de F un peptide de fusion qui s'insère dans la membrane-cible durant le processus de fusion (Lamb and Kolakofsky, 2001). Après fusion, les RNP viraux sont relargués dans le cytoplasme, où il est probable qu'ils interagissent avec des composants du cytosquelette cellulaire (pour une revue, voir (Lecellier *et coll.*, 2002)).

Ces dernières interactions ainsi que les étapes initiales de la réplication sont très mal connues.

TRANSCRIPTION ET REPLICATION

Transcription

Après l'entrée et le relargage des RNP dans le cytoplasme, et en réponse à un signal inconnu, les polymérases virales présentes sur les RNP commencent la transcription séquentielle des ARNm viraux. La première étape dans la transcription est la synthèse d'une molécule d'ARN leader. Après avoir transcrit ce dernier, la polymérase s'arrête à la séquence de terminaison du leader et libère l'ARN leader. Elle a alors deux choix : se décrocher de la matrice et donc arrêter la transcription, ou bien ré-initier la synthèse de l'ARNm suivant (celui de N) après la courte séquence intergénique. De même, à chaque jonction, après avoir libéré l'ARNm qu'elle transcrivait, la polymérase a le choix entre quitter la matrice ou ré-initier la synthèse de l'ARNm suivant. La fréquence de ré-initiation à chaque jonction est élevée mais inférieure à 100%, ce qui génère un gradient dans la quantité des différents ARNm synthétisés. Ce gradient décroissant suit l'ordre des gènes, et le dernier ARNm qui spécifie l'enzyme L est donc le moins synthétisé (Ray and Fujinami, 1987). A la différence de l'ARN leader, qui possède une extrémité 5' triphosphate (Curran and Kolakofsky, 1999), les ARNm sont coiffés en 5' et polyadénylés en 3'. Notre compréhension de ce phénomène et du rôle des séquences d'initiation ou de terminaison est amenée à beaucoup progresser grâce à l'utilisation de virus modifiés par génétique inverse (Marriott and Easton, 1999; Nagai, 1999). Ainsi, des mutations de la séquence d'initiation du gène G de VSV, un *Rhabdoviridae*, entraînent la synthèse d'ARNm qui sont transcrits en nombre inchangé, mais ne sont pas coiffés et se terminent de façon précoce (Stillman and Whitt, 1999). Ceci suggère que les événements d'initiation et de coiffage des transcrits sont séparés.

Réplication

A un certain moment du cycle viral commence la réplication de l'ARN génomique, toujours en réponse à un signal inconnu. La polymérase virale, dans ce cas, ne s'arrête plus à la séquence de terminaison du leader, mais continue la synthèse ininterrompue de l'ARN génomique, qui est encapsidé de façon parallèle (voir figure 2). On sait que la synthèse d'ARN et son encapsidation sont étroitement liés (Gubbay *et coll.*, 2001), mais le mécanisme exact de ce couplage reste encore à définir.

Transition entre transcription et réplication

Quels événements déterminent le passage de la polymérase en mode réplicatif ? Tout d'abord, dans la phase précoce de l'infection, avant que les produits de traduction viraux ne se soient accumulés, la polymérase virale se borne à transcrire l'ARN leader et les ARNm viraux. Ce n'est qu'avec l'accumulation des protéines virales, en particulier, N, P, et L, que l'encapsidation peut matériellement commencer. Un modèle simple et répandu, fondé sur une analogie avec le VSV, postule que c'est le degré d'abondance de N qui gouverne la transition entre transcription et réplication, par un mécanisme inconnu. Cette hypothèse est simple et séduisante, mais en l'absence de presque tout support expérimental, elle doit n'être considérée comme une hypothèse parmi d'autres. Il existe toutefois au moins une observation qui rentre dans le cadre de cette théorie. La polymérase du virus de la rougeole synthétise à un taux très faible un ARN bicistronique (leader-N) encapsidé et polyadénylé (Castaneda and Wong, 1990). Cet ARN serait produit par une polymérase en mode réplicatif qui aurait débuté l'encapsidation du leader, et donc ignoré la jonction leader-N, puis qui serait repassée en mode transcriptionnel par manque de substrat N°. Cette polymérase aurait alors polyadénylé et relargué l'ARN néosynthétisé.

Le mécanisme présumé reliant l'encapsidation de l'ARN lors de la réplication avec le fait que la polymérase passe (ou reste) en mode réplicatif n'est pas clair. Une première question à élucider serait de savoir si c'est la *polymérase* qui assemble N sur l'ARN lors de l'élongation (on peut alors penser que les deux procédés sont étroitement couplés), ou si N s'auto-polymérise *tout seul* sur l'ARN néosynthétisé, comme elle en a la capacité (il y aurait sans doute alors un délai entre synthèse de l'ARN et encapsidation, représenté par exemple en figure 2). Cette question est évidemment indissociable d'un autre problème : est-ce le complexe N°-P qui (par son abondance ?) fait passer la polymérase en mode réplicatif, ou est-ce au contraire la polymérase qui rend N°-P compétent pour l'encapsidation ? A ce sujet, une observation intéressante concerne la souche Z de SeV. Celle-ci possède la particularité intéressante que la polymérase traverse la jonction leader-N avec une haute fréquence (20%). Cette particularité a pu être attribuée à une substitution dans la région N_{TAIL} de SeV, indiquant qu'une interaction entre la polymérase virale et N_{TAIL} régule ce phénomène, et confirmant le rôle actif de N dans la synthèse d'ARN (J. Curran, communication personnelle).

Enfin, une expérience très étonnante de réplication hétérologue pourrait jeter une lumière inattendue sur la réplication (Chandrika *et coll.*, 1995). La protéine N du virus de la rougeole pourrait (en l'absence de P du virus de la rougeole) encapsider un ARN génomique de SeV, produit par la polymérase de SeV à partir d'un RNP de SeV. Le produit d'encapsidation est sensible aux nucléases, ce qui indique que l'encapsidation est incorrecte. De plus, les auteurs pensent que c'est la forme *déjà assemblée* de N qui aurait encapsidé l'ARN génomique. Cette expérience (jamais reproduite) hautement inattendue

pourrait, si elle est exacte, recéler plusieurs clés de la réplication. En premier lieu, comment est-il possible que N^{NUC} puisse encapsider un ARN ? On pourrait penser qu'"encapsider" n'est pas le terme exact, et que par exemple, l'ARN est lié à l'extérieur de N^{NUC}. Par ailleurs, dans la même expérience, N de SeV seul (c'est-à-dire en l'absence du P homologue) ne peut pas servir de substrat à l'encapsidation. On pourrait donc penser que L et/ou P de SeV "empêchent" la nucléoprotéine homologue, mais pas celle d'un autre virus, d'être compétente pour la réplication (par exemple en phosphorylant N ?) …
Cette expérience mérite en tout cas certainement d'être clarifiée !

ASSEMBLAGE ET BOURGEONNEMENT

Les mécanismes par lesquels la particule virale est assemblée à la membrane sont inconnus. Ils semblerait qu'ils soient assez différents entre les différents genres des *Paramyxovirinae*. On pense que M, qui possède la capacité de s'auto-assembler serait en quelque sorte "le chef d'orchestre" de l'assemblage du virus (Bohn *et coll.*, 1990; Mottet *et coll.*, 1999). En particulier, les RNP sont incorporés dans le virion grâce à des interactions entre N et M (Coronel *et coll.*, 2001; Stricker *et coll.*, 1994).
Les interactions entre protéines impliquées dans l'assemblage du virion doivent être spécifiques, puisque les protéines de membrane cellulaires sont en majorité exclues des virions (Lamb and Kolakofsky, 2001). On pense que M interagit avec les queues cytoplasmiques des glycoprotéines F et H. Takimoto *et coll.* n'ont pu démontrer de telles interactions directes (Takimoto *et coll.*, 2001). Dans le cas du virus de la rougeole, les micro-domaines enrichis en glycosphingolipides et en cholestérol pourraient permettre la co-localisation des protéines virales impliquées dans l'assemblage (Manie *et coll.*, 2000; Vincent *et coll.*, 2000).

ROLE DE PROTEINES CELLULAIRES DANS LE CYCLE REPLICATIF

Le rôle d'éventuelles protéines cellulaires dans la réplication virale est très mal connu. On sait notamment que 3 protéines cellulaires se lient aux régions leader 3' et 5' de l'ARN du virus de la rougeole, mais elles n'ont pu être identifiées (Leopardi *et coll.*, 1993). Le rôle de protéines du cytosquelette telles que l'actine ou la tubuline dans la réplication des *Mononegavirales* est connu depuis longtemps (Choudhary *et coll.*, 2000; De and Banerjee, 1999; Moyer *et coll.*, 1990) ; en particulier, la tubuline serait associée à L dans le cas du virus de la rougeole (Moyer *et coll.*, 1990) (pour une revue sur le rôle des protéines du cytosquelette dans le cycle viral, voir (Lecellier *et coll.*, 2002)). La figure 6 illustre bien la nécessité pour la synthèse de complexes macromoléculaires aussi gigantesques que les RNP (formés d'environ 3000 protéines (Lamb and Choppin, 1977a) !) d'un ancrage sur le cytosquelette.

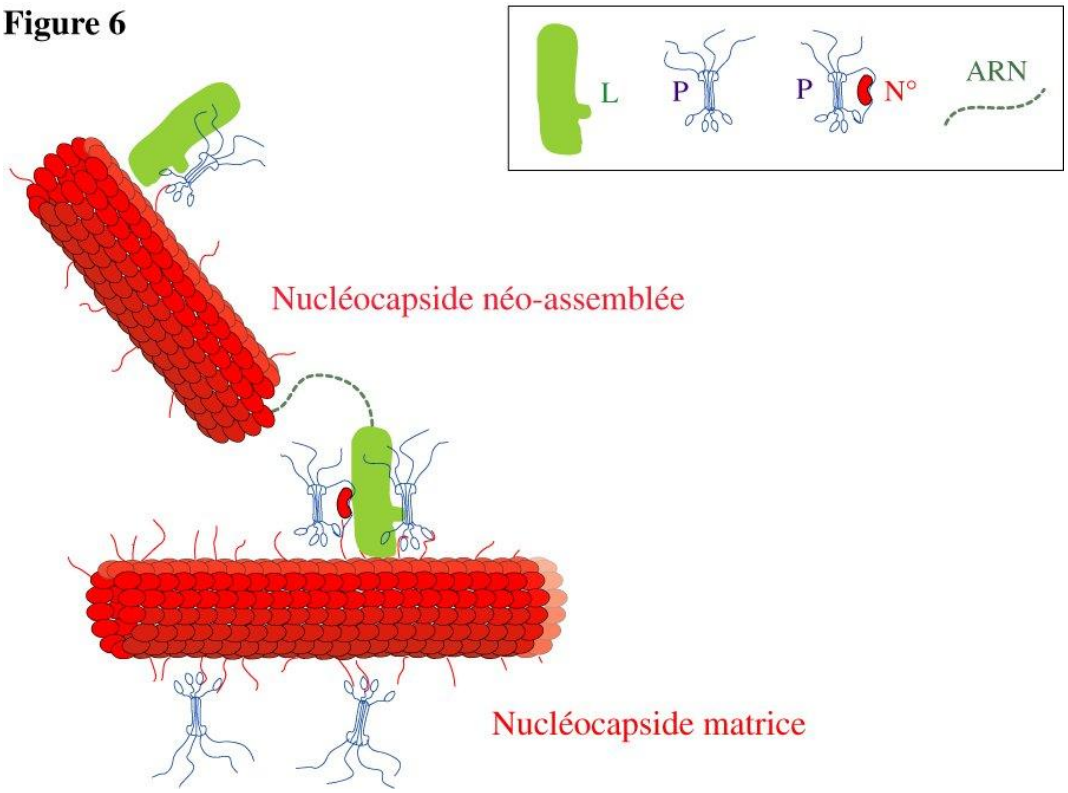

Figure 6. Les nucléocapsides virales doivent être ancrées au cytosquelette.

Vue d'artiste d'une polymérase en train de répliquer l'ARN viral, qui est encapsidé de concert. Les conventions sont explicitées dans la figure. Cette figure souligne la nécessité pour les nucléocapsides virales (y compris la nucléocapside néo-assemblée) d'être ancrées au cytosquelette. En effet les nucléocapsides (qui mesurent 1 μm de long pour 20 nm de diamètre) sont des objets microscopiques soumis à des forces de viscosité (au contraire d'objets beaucoup plus petits comme les tétramères de P, ou L). Ces forces de viscosité arracheraient la polymérase virale de sa matrice si tous les partenaires de la réplication n'étaient pas ancrés !

Si on admet ce postulat, un problème majeur apparaît immédiatement : comment la réplication peut-elle avoir lieu dès lors que la nucléocapside matrice et la nucléocapside néo-assemblée sont *toutes deux* ancrées au cytosquelette ? La nucléocapside matrice pourrait tourner sur elle-même, pendant que la polymérase progresse linéairement ; l'ensemble (polymérase+nucléocapside néo-assemblée) pourrait tourner autour de la nucléocapside matrice ; ou par exemple, la polymérase pourrait tourner seule autour de la nucléocapside matrice et couper puis rabouter l'ARN néosynthétisé à chaque tour.

Toutes ces hypothèses supposent que la nucléocapside matrice et la nucléocapside néo-assemblée conservent un axe parallèle lors de la réplication, sans doute contraint par les fibres du cytosquelette.

La formation de **corps d'inclusion cytoplasmiques** lors de l'infection par les *Paramyxovirinae*, correspondant sans doute aux sites de la réplication et de la transcription virale, pourrait refléter l'interaction étroite du complexe de réplication viral et du cytosquelette lors de ces processus.

Par ailleurs la figure pose un problème intéressant : les polymérases commencent-elles à transcrire immédiatement l'ARN en cours d'encapsidation ? (La figure montre une polymérase sur la nucléocapside néo-assemblée). Sinon, pourquoi ?

Par ailleurs, on sait que la réplication est stimulée par la protéine de choc thermique hsp70, une piste de recherche explorée depuis près de 20 ans par le groupe de Oglesbee ; ainsi, un choc thermique permet d'améliorer la production de virus par génétique inverse (Oglesbee *et coll.*, 1996; Parks *et coll.*, 1999), ce qui est bien corrélé avec la présence de cette protéine sur les nucléocapsides de *Morbillivirus*, au sein de cellules infectées (Oglesbee *et coll.*, 1990).

QUELQUES CHERCHEURS ET LEURS METHODES D'INVESTIGATION

Certains chercheurs spécialisés dans l'étude du complexe de réplication des *Paramyxoviridae* utilisent des méthodes particulièrement simples, robustes, et qui ont fait leurs preuves. Ces méthodes peuvent et doivent être une source d'inspiration. Il m'a paru intéressant de les mentionner ici.

- Depuis 1996, le groupe de Y. Ito a mené la cartographie la plus complète des interactions N-P-L sur un *Paramyxoviridae*, le *Rubulavirus* hPIV-2. Dans ce dessein, il a généré en tout plus de 80 **anticorps monoclonaux** contre N et P. Après avoir déterminé les sites de reconnaissance des anticorps sur N et P, son groupe a conduit des études de co-immunoprécipitation sur les complexes formés par des mutants de délétion de N, P, ou L. Il a ainsi cartographié les régions de N, P et V impliquées dans les interactions N-N, N-P, N-V, N-L, et P-L (Nishio *et coll.*, 2000; Nishio *et coll.*, 1999; Nishio *et coll.*, 1997; Nishio *et coll.*, 1996; Watanabe *et coll.*, 1996). En toute logique, leur prochain article devrait présenter la cartographie des régions de L impliquées dans l'interaction L-P grâce à des anticorps anti-L … Une telle approche leur permettra aussi à terme de cartographier les régions fonctionnelles de ces protéines, en inhibant la transcription ou la réplication au moyen d'anticorps (voir par ex. (Deshpande and Portner, 1984; Gill *et coll.*, 1988)).

- Le groupe de S. Moyer applique avec constance et succès une stratégie de **substitution de résidus conservés, chargés ou polaires, par des alanines**. La logique sous-jacente, amplement vérifiée, est que ceux-là sont souvent impliqués dans des interactions protéine-protéine ou dans des réactions enzymatiques. Leur substitution par des alanines perturbe en général peu la structure de la protéine (Bowman *et coll.*, 1999; Cortese *et coll.*, 2000; Feller *et coll.*, 2000; Grogan and Moyer, 2001; Holmes and Moyer, 2002; Myers and Moyer, 1997; Myers *et coll.*, 1999; Smallwood *et coll.*, 1999; Tuckis *et coll.*, 2002). Un article récent suggère qu'une substitution par des sérines serait encore moins délétère (Jonson and Petersen, 2001). Chacun de ces *9 articles* a permis de différencier, voire de découvrir, plusieurs fonctions de L ou de P.

- La **substitution** d'une ou plusieurs protéines virales par les **protéines d'un virus homologue**, ou de façon plus fine, d'une *région* d'une protéine virale par la région homologue d'un virus proche peut apporter des informations précieuses. Par exemple, une telle substitution permet de définir quelles protéines sont impliquées dans des interactions, ou de définir plus précisément les régions d'interaction. J. Curran *et coll.* ont appliqué cette méthode avec succès, (Curran *et coll.*, 1993; Curran *et coll.*, 1994; Curran and Kolakofsky, 1991) ; citons aussi (Bousse *et coll.*, 2001).

LE DESORDRE STRUCTURAL DANS LES PROTEINES : UNE REFONTE DE NOTRE CONCEPT D'UNE STRUCTURE TRIDIMENSIONNELLE INTANGIBLE

L'interprétation de la fonction d'une protéine en termes de structure tridimensionnelle est au cœur de notre appréhension de la biochimie. Le fait que la fonction repose sur la structure est même inscrite dans notre langage : nous parlons en jargon d'études "structure-fonction", et les mots "protéines non structurées" et "protéines dénaturées" sont utilisés de façon interchangeable. De plus, l'avalanche de structures tridimensionnelles de protéines déterminées par diffraction aux rayons X et par RMN a occupé le devant de la scène ces dernières années. La résolution des structures de toutes les protéines existantes (non homologues) fait l'objet de projets à très grande échelle et disposant donc de très gros moyens. Nous n'avons pas prêté attention aux nombreux contre-exemples s'accumulant au cours des années, qui indiquaient l'existence de fonctions ne reposant pas sur une structure fixe. Le premier appel à la reconnaissance de protéines contenant des régions désordonnées (c'est-à-dire non structurées) est très récent (Wright and Dyson, 1999). Les auteurs suggéraient que l'existence de protéines désordonnées exigeait une refonte du paradigme que la structure gouverne la fonction dans les protéines. Cet appel a été suivi par un grand nombre de communications. Des protéines dont l'étude avait été abandonnée (parce qu'on pensait qu'il devait y avoir "un problème" pour qu'elles n'aient pas de structure secondaire) ont refait surface (K. Dunker, communication personnelle). Pourtant, les cristallographes connaissaient bien le désordre et la flexibilité dans les protéines. La plupart des structures tridimensionnelles qu'il résolvent comportent des régions de densité électronique manquante, signalant une mobilité des résidus correspondants. Certes, il est rare qu'une région longue (de plus de 30 aa) soit invisible - notamment parce que des longues régions désordonnées empêchent la cristallisation ! - mais de nombreux exemples existent et sont largement connus. De plus, de nombreuses protéines résistent à toutes les tentatives de cristallisation tant qu'on ne leur adjoint pas un partenaire (ARN, ligand, anticorps), ce que les cristallographes eux-mêmes interprètent en termes de flexibilité ou de désordre structural. Pourquoi alors ce désintérêt des protéines non structurées ? D'une part, les cristallographes sont amenés à penser en termes de "cristallisation à tout prix", car il s'agit d'une procédure longue et fastidieuse. D'autre part, on ne conçoit bien que ce qu'on peut se représenter. La structure tridimensionnelle d'une protéine est tangible, esthétique, et surtout fait énormément progresser notre compréhension des mécanismes par lesquels elle exerce son action. En revanche, que comprend-on d'une absence de structure, que par

définition il est difficile de se représenter ? On trouvera une discussion de ce point dans le commentaire de la revue, à la fin de la thèse.
Les enzymologistes sont aussi très réticents à concevoir l'absence de structure, tant ils sont familiers de la beauté des mécanismes délicats des enzymes, gouvernés par la structure précise de ces véritables machines-outils perfectionnées. Pourtant, nous pensons que cette thèse prouve que des machineries enzymatiques aussi extraordinairement sophistiquées que celles des *Paramyxovirinae* peuvent faire un usage abondant du désordre structural.
Si la séquence en acides aminés détermine la structure tridimensionnelle, elle devrait aussi déterminer l'absence de structure tridimensionnelle. Les études de K. Dunker *et coll.*, qui ont développé un logiciel de prédiction des régions désordonnées des protéines sur la base de leur séquence, en fournissent la preuve (Dunker *et coll.*, 2001). Dans ce cadre, on s'étonne que les projets de résolution de structures à haut débit, mentionnés plus haut, aient souvent sous-estimé la fréquence de régions désordonnées dans les protéines (estimée à plus de 60% pour certains eucaryotes ! (Dunker and Obradovic, 2001)), et continuent à le faire.

Il n'est pas notre propos de décrire ici en détail les très nombreux exemples, souvent fascinants, de rôles du désordre structural. On trouvera une revue claire et exhaustive (tant d'un point de vue scientifique qu'historique) dans (Dunker *et coll.*, 2001). Nous souhaiterions cependant souligner un point étonnant. Les organismes multi-cellulaires semblent contenir beaucoup plus de protéines ayant des régions désordonnées que les organismes unicellulaires (Dunker and Obradovic, 2001). La plupart des protéines désordonnées décrites dans cette dernière étude occupent des fonctions de signalisation ou de régulation – on conçoit que la coordination des organismes multi-cellulaires fasse appel à un nombre plus important de protéines régulatrices. Sans doute n'est ce pas une coïncidence que des protéines désordonnées soient impliquées dans des interactions avec de multiples partenaires ; c'est en particulier le cas des protéines N et P des *Paramyxovirinae*, qui font l'objet de cette thèse.

Cette revisite de nos modes de pensées en termes de structure tridimensionnelle établie a des implications jusqu'au cœur de la recherche en pharmacologie, un bastion traditionnel du paradigme structure-fonction. La conception rationnelle d'agents médicamenteux repose sur une connaissance de la structure tridimensionnelle de la cible du médicament, conçu pour se lier à sa cible par un mécanisme de type "clé-serrure". Pourtant, des polypeptides cationiques anti-microbiens jouent un rôle majeur dans l'immunité innée et adaptative. Or ces polypeptides sont non structurés et linéaires, ou bien se structurent grâce à des ponts disulfure (c'est le cas de la famille des défensines, qui revêt une importance capitale dans la défense de l'hôte) (Raj and Dentino, 2002). Ainsi, pour certaines protéines, une structure tridimensionnelle fixe n'est pas nécessaire à une activité cruciale à notre survie.

ARTICLE N°1 : "THE N-TERMINAL DOMAIN OF THE PHOSPHOPROTEIN OF *MORBILLIVIRUSES* BELONGS TO THE NATIVELY UNFOLDED CLASS OF PROTEINS"

RESUME

Le domaine N-terminal de la phosphoprotéine du virus de la rougeole est intrinsèquement désordonné.

Nous avons exprimé, purifié, et caractérisé le domaine N-terminal (PNT) de la phosphoprotéine du virus de la rougeole. À l'aide des techniques de résonance magnétique et nucléaire (RMN), dichroïsme circulaire, gel-filtration et diffusion dynamique de la lumière, nous montrons que PNT n'est pas structuré en solution. Deux approches bio-informatiques complémentaires confirment que PNT appartient à la classe récemment décrite des protéines non structurées à l'état natif, dites "intrinsèquement désordonnées". Nous étendons ces résultats aux domaines N-terminaux d'autres phosphoprotéines de *Morbillivirus*, et à la protéine W du virus de Sendai, un *Respirovirus*. Certaines protéines intrinsèquement désordonnées se structurent partiellement lors de l'interaction avec leurs partenaires, un processus appelé "repliement induit". Par protéolyse en présence de trifluoroéthanol (TFE), nous avons identifié un élément de structure secondaire qui pourrait se structurer lors d'un tel repliement induit.

INTRODUCTION

La phosphoprotéine des *Paramyxovirinae* est une protéine modulaire qui possède deux parties fonctionnellement et structuralement distinctes : la partie C-terminale (PCT), qui a fait l'objet du plus grand nombre d'études, et la partie N-terminale (PNT) qui n'avait pas, jusqu'à cette étude, été caractérisée d'un point de vue structural.

À l'origine notre but était de produire la phosphoprotéine en entier en vue de sa cristallisation. Cependant, des problèmes d'agrégation et de protéolyse (que nous avons ultérieurement attribué à la partie C-terminale de P) nous ont conduit à exprimer et à purifier séparément les deux modules de P. Nous avons constaté rapidement que le domaine N-terminal, qui se purifiait aisément et apparaissait homogène en gel-filtration, ne formait pas de cristaux dans un grand nombre de conditions standard. Nous l'avons alors soumis à une analyse par résonance magnétique nucléaire, qui a montré que PNT ne semblait pas contenir de structure secondaire constante. Un article de recherche très récemment paru (Uversky *et coll.*, 2000a) a alors attiré notre attention. Celui-ci montrait que des attributs de

séquence très simples permettaient de prédire avec une grande fiabilité le fait qu'une protéine soit intrinsèquement désordonnée. Les attributs de séquence de PNT correspondaient à ceux d'une protéine intrinsèquement désordonnée. De plus, un logiciel de prédiction des régions désordonnées des protéines, de conception récente, PONDR, (Li *et coll.*, 1999), aboutissait au même résultat. Nous avons alors entrepris une étude structurale plus fine de PNT, qui fait l'objet du présent article, et qui a montré que PNT possédait une conformation étendue en solution. Le fait que PNT soit intrinsèquement désordonné a de nombreuses implications, en partie discutées dans la revue : d'une part, il est bien sûr impossible de le cristalliser en l'état (en l'absence de tout partenaire). D'autre part, sa conformation étendue et variable pourrait favoriser la formation de complexes avec de nombreuses protéines de formes tridimensionnelles différentes, peut-être simultanément, une hypothèse reprise dans la revue qui suit. Enfin, PNT est susceptible de s'étendre sur une grande distance, ce qui pourrait être nécessaire à l'encapsidation de l'ARN néosynthétisé par N°.

The N-terminal domain of the Phosphoprotein of Morbilliviruses belongs to the natively unfolded class of proteins

David Karlin, Sonia Longhi, Véronique Receveur, and Bruno Canard*

Running title: Measles virus PNT is natively unfolded

Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Université Aix-Marseille I et II, ESIL, Campus de Luminy, 13288 Marseille Cedex 09, France

(*) to whom requests for reprints should be addressed
ESIL-CNRS-AFMB Case 925
163, avenue de Luminy
13288 Marseille Cedex 09
France
E-mail: **bruno@esil.univ-mrs.fr**
Tel: (33) (491) 82 86 44
Fax: (33) (491) 82 86 46

Sous presse dans *Virology*

SUMMARY

We report the bacterial expression, purification, and characterization of the N-terminal domain (PNT) of the measles virus phosphoprotein. Using nuclear magnetic resonance, circular dichroism, gel filtration and light scattering, we show that PNT is not structured in solution. We show by two complementary computational approaches that PNT belongs to the recently described class of natively unfolded proteins, further confirming its reported similarity with acidic activation domains of cellular transcription factors. We extend these results to the N-terminal domains of other *Morbillivirus* Phosphoproteins and to the corresponding protein W of Sendai virus, a *Paramyxovirus*. Unstructured proteins may undergo some degree of folding upon binding to their partners, a process termed «induced folding». Using limited proteolysis in the presence of trifluorethanol (TFE), we identified residues 27 to 38 as a putative secondary structure element of PNT arising upon induced folding.

KEY WORDS

measles virus, sendai virus, natively unfolded, unstructured protein, acidic activation domain, trifluoroethanol, phosphoprotein, rinderpest virus, paramyxoviridae, morbillivirus

INTRODUCTION

Measles virus (MV), a member of the family *Paramyxoviridae*, is a major human pathogen, responsible annually for more than one million deaths worldwide. Despite the clinical importance of MV and the recent successes of antiviral therapies directed against viral polymerases, very little is known about the MV replication machinery. MV is an enveloped virus with a non-segmented, single stranded, negative RNA genome. The genomic RNA is packaged by the Nucleoprotein (N) to form the viral nucleocapsid. Transcription and replication are carried out on this (N:RNA) complex by the polymerase L in association with the Phosphoprotein P. In the last decade, much has been elucidated about the multiple roles played by the key protein P in conjunction with N and L (Curran and Kolakofsky, 1999). P and N each perform several functions during the viral cycle. Both transcription and replication take place onto the assembled form of N (N^{NUC}), but the replication step also necessitates the concomitant assembly of the soluble, unassembled form of N (N°), onto the newly synthesized viral genomic RNA, with the participation of P. At least in the case of MV, N° undergoes a conformational maturation before being encapsidated (Gombart, Hirano, and Wong, 1993). In the current state of our knowledge, it seems clear that P plays a pivotal role together with N in the transition from transcription to replication. P is a modular protein with 2 distinct functional domains (Lamb and Kolakofsky, 2001). The P C-Terminal domain (PCT), which is the most conserved in sequence, contains all the regions required for viral transcription. On the other hand, the poorly conserved N-Terminal domain (PNT) provides several additional functions required for replication (Curran, 1998; Curran and Kolakofsky, 1999). PNT is also distinguished by containing most phosphorylation sites of P. Although phosphorylation of *Paramyxoviridae* phosphoproteins is highly conserved, implying functional significance, its role is still unclear (Lamb and Kolakofsky, 2001).

All *Paramyxovirus* P proteins form oligomers through a coiled-coil motif located within PCT (Curran et al., 1995). The recent structure determination of this oligomerization domain shows that it is tetrameric (Tarbouriech et al., 2000), suggesting that P may form tetramers rather than trimers as previously proposed (Curran et al., 1995). These very stable oligomers are the active form of P for replication and transcription (Curran, 1998). PCT also contains the region responsible for binding to L (Smallwood, Ryan, and Moyer, 1994) and those necessary for binding to the assembled form of N, N^{NUC} (Harty and Palese, 1995; Ryan, Morgan, and Portner, 1991). The N-terminal part of P, PNT, is not characterized from a structural point of view. Remarkably, PNT is found as a domain in at least two viral proteins, P and V, in all *Morbilliviruses* and *Rubulaviruses* (see below) and in some *Paramyxoviruses*. V originates from a transcriptional editing event occuring on approximately 50 % of P mRNAs. An extra G nucleotide is inserted at position 753 of the mRNA of MV P, and translation of the edited RNA proceeds in the reading frame –1 to yield a new protein, V

(Cattaneo et al., 1989) (see Figure 1). The 230 residues shared by the amino-termini of MV P and V constitute the N-terminal domain of P, PNT. The 69 highly conserved C-terminal residues of V, unique to the V protein, form a zinc-binding domain (Liston and Briedis, 1994). A similar editing mechanism yields SeV V protein but also a third protein, W, which is composed solely of the N-terminal domain of P (Figure 1). Thus the PNT domain of SeV is expressed on its own as the protein W (Lamb and Kolakofsky, 2001).

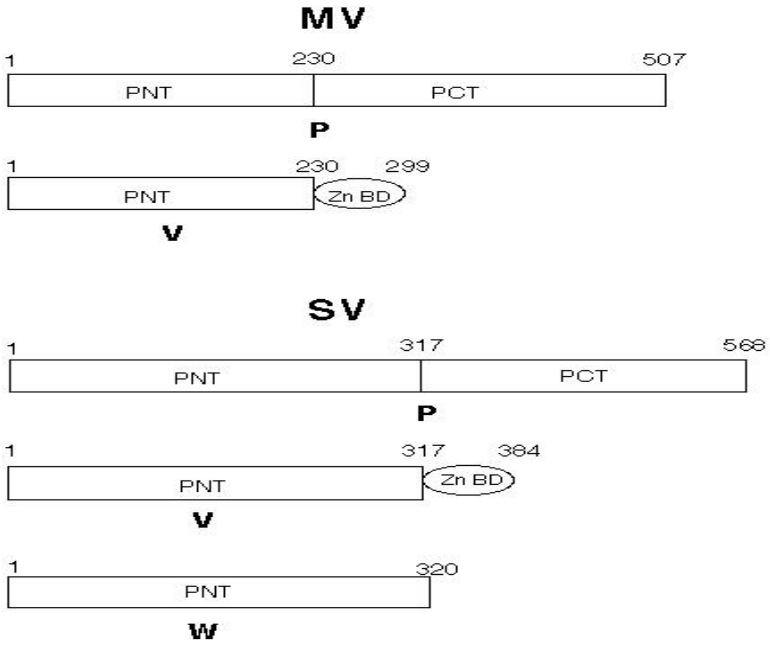

Figure 1. Domain organization of the P and V proteins of MV and SeV

The Phosphoprotein N-Terminal (PNT) domain common to P and V, the Phosphoprotein C-Terminal (PCT) domain unique to P, and the Zinc Binding Domain (Zn BD) unique to V are shown. In the case of SeV, the PNT domain is also expressed alone, as a protein, W.

Auxiliary MV V protein as well as SeV V and W proteins have a dual nature. On one hand, they have an inhibitory effect on viral replication (Horikami, Smallwood, and Moyer, 1996; Tober et al., 1998) and they are dispensable for replication in cultured cells (Delenda et al., 1997; Schneider, Kaelin, and Billeter, 1997). On the other hand, they are necessary for virulence in mice (Delenda et al., 1998; Patterson et al., 2000). The pathogenicity determinants of V lie in its C-terminal zinc-binding domain (Kato et al., 1997). Conversely, the inhibitory effect of V on replication has been mapped to its N-terminal domain PNT, by interfering with encapsidation of the nascent RNA (Horikami, Smallwood, and Moyer, 1996; Tober et al., 1998). Finally, V interacts with the polymerase L via the PNT domain as shown for Rinderpest Virus (RPV), closely related to MV (Sweetman, Miskin, and Baron, 2001).

As mentioned above, although the N-terminal part of P is dispensable for transcription, it is required for replication. Elegant studies by Curran *et al.* (Curran, Marq, and Kolakofsky, 1995; Curran, Pelet, and Kolakofsky, 1994) have shown that SeV PNT domain acts like a

chaperone for N, preventing its illegitimate self-assembly in the absence of genomic RNA synthesis. Rather, N is kept in a soluble form N° by directly interacting with PNT; this chaperone activity of P has been precisely mapped to aa 33-41. N°-P binding also necessitates the last C-terminal 90 residues of P. Furthermore, two regions (aa 1-77 and 78-156) of PNT that are essential for genomic RNA synthesis have been uncovered. These regions are reminiscent of the acidic activation domains (AADs) of cellular transcriptional factors in that they seem to be position-independent and can partially substitute each other. Other studies suggest that *Morbillivirus* PNT plays a role similar to that of SeV PNT in keeping N° soluble by binding to it (Curran, Marq, and Kolakofsky, 1995; Curran, Pelet, and Kolakofsky, 1994; Harty and Palese, 1995; Shaji and Shaila, 1999; Tober et al., 1998).

The N-terminal domains of *Rubulavirus* P and V proteins have distinct properties from those of *Paramyxoviruses* and *Morbilliviruses*. They are shorter and of basic nature; furthermore the V protein of *Rubulaviruses* is encapsidated in virions, contrary to that of *Paramyxoviruses* and *Morbilliviruses* (Delenda et al., 1997; Lamb et Kolakofsky, 2001). Although they are related (functionally and in sequence) to the set of C', C, Y1, Y2 proteins of *Paramyxoviruses* (Lamb and Kolakofsky, 2001), they may share at least one function with *Morbillivirus* and *Paramyxovirus* PNT, since SV5 PNT might have a role in keeping N° soluble prior to encapsidation (Precious et al., 1995).

As a first step in studying the role of P in viral replication, we report the expression, purification and characterization of PNT. We found that PNT is unstructured in aqueous solution, and identified a region within PNT that has the propensity to undergo an unstructured / structured transition under hydrophobic conditions.

RESULTS

P is a modular protein, divided into two modules: a C-terminal half (PCT) sufficient for viral transcription and an N-terminal half (PNT) that performs several additional functions in replication. In order to investigate in detail the different functions of P in relation with its structure, the N-terminal domain of P was isolated and characterized.

Domain analysis of P and subcloning of PNT

The modular organization of the P gene reviewed above and described in Figure 1 suggests that residues 1-230 constitute an independent functional domain, PNT. We have cloned the DNA fragment of the P gene encoding PNT (residues 1-230) into the pET21a expression vector with an N-terminal hexahistidine tag in order to facilitate purification. The corresponding plasmid is referred to as pET21a/$PNT_{H6.}$

Expression and purification

The P gene of MV contains numerous arginine codons AGG that are used with a very low frequency in *E. coli*. Accordingly, co-expression of the plasmid pUBS520 (Brinkmann, Mattes, and Buckel, 1989), which supplies the corresponding rare tRNA in *trans* to pET21a***/***PNT_{H6} was required for expression of PNT (not shown). Most PNT produced was recovered from the soluble fraction of bacterial lysates (Figure 2, lane 2). PNT was purified to homogeneity (> 95%) in 3 steps: Immobilized Metal Affinity Chromatography (IMAC), ion exchange chromatography and gel filtration (Figure 2, lanes 3-5). The identity of PNT was confirmed using immunodetection (Western blotting) with an anti-hexahistidine monoclonal antibody (not shown). As shown in Figure 2, PNT displays an abnormally slow migration in SDS-PAGE with an apparent molecular weight (MW) of 33 kDa (expected weight 25 kDa). This abnormal behavior has been reported previously for all P proteins of *Morbilliviruses* and *Paramyxoviruses* (Lamb and Kolakofsky, 2001) and attributed to the occurrence of acidic stretches in PNT, which is consistent with our result. The theoretical MW of PNT (24.823 kDa) was confirmed by mass spectrometry.

In order to characterize the structural properties of PNT, we performed on this domain biophysical and biochemical studies allowing to analyze its secondary and tertiary structure.

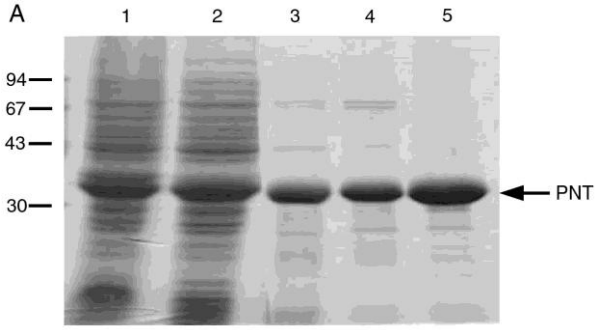

Figure 2. Purification of PNT

12% SDS-PAGE. Lane 1: Bacterial lysate (total fraction). Lane 2: Clarified supernatant (soluble fraction). Lane 3: Eluent from IMAC. Lane 4 : Eluent from Ion Exchange Chromatography. Lane 5 : Eluate from Gel Filtration.

Nuclear Magnetic resonance (NMR) experiments

The structure of PNT was first investigated using 2D NMR. Figure 3 shows the amide region of a NOESY spectrum. The very small spread of the resonance frequencies for amide protons (between 7.8 ppm and 8.7 ppm, see frame in Figure 3) together with the absence of NOEs in the amide-amide region are typical of a protein without any constant secondary structure. Moreover, the absence of any specific short-range interactions as those required for stabilizing an oligomer, suggests that the protein is monomeric.

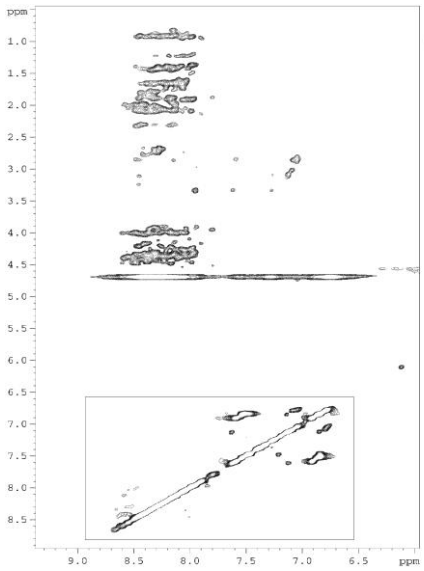

Figure 3. NMR spectrum of PNT

Two-dimensional 1H-NMR spectrum of a 0.3 mM solution of purified PNT at 283 K. ppm quotes for resonance shifts in part per million of the spectrophotometer frequency. The frame shows the small spread of the resonance frequencies for amide protons.

Circular dichroism (CD) studies

The CD spectrum of PNT at neutral pH lacks evidence of any significant ordered secondary structure, as seen from its large negative ellipticity at 198 nm and nil ellipticity at 185 nm (Figure 4).

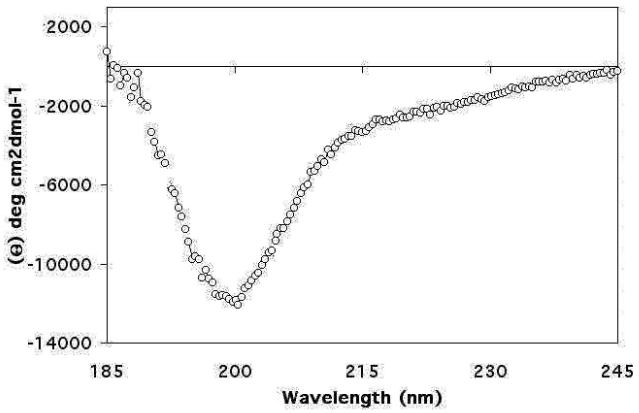

Figure 4. CD spectrum of PNT

The spectrum was recorded in 10 mM sodium phosphate buffer, pH 7 (see Materials and Methods).

Gel filtration

PNT elutes from the gel filtration column as a sharp peak indicating the presence of a well-defined species. This peak corresponds to a 115 kDa globular protein (Figure 5). The elution volume of a protein from a gel filtration column depends in fact, not on its MW, rather on its hydrodynamic properties. The hydrodynamic radius of a protein (Stokes radius, R_s) can be deduced from its apparent MW (as seen by gel filtration) (Uversky, 1993). The apparent MW of 115 kDa measured for PNT corresponds to a Stokes radius of 41 $\pm$ 3 Å . In addition, the theoretical Stokes radius of a monomeric, native (R_sN) or unfolded (R_sU) protein can be estimated according to the equations described in (Uversky, 1993). In the case of PNT, R_sN=23 Å and R_sU=46 Å. Therefore, the Stokes radius measured for PNT (41 Å) is not compatible with a monomeric, globular protein. Such a large value of the Stokes radius could be attributed either to a high degree of multimerization or to a monomeric, unfolded protein (R_sU predicted: 46 Å). Multimers in the form of non specific aggregates would have led to a broad peak in gel filtration, which is not the case here. On the other hand, the hypothesis of an unfolded protein is consistent with the CD and NMR data, which show that PNT lacks stable secondary structure.

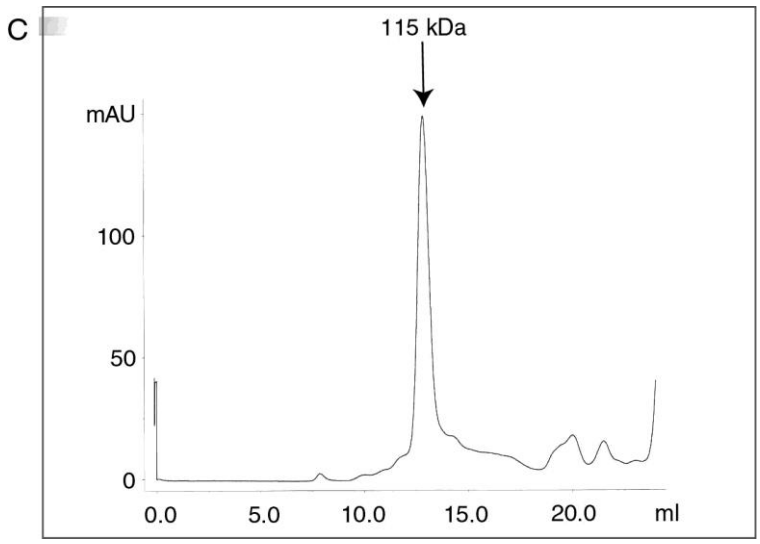

Figure 5. Gel filtration of PNT onto a Superdex 200 column

Elution was carried out in 50 mM sodium phosphate buffer pH 7, 100 mM NaCl. The main peak containing PNT is highlighted by an arrow and the mass deduced from the column calibration is shown.

Dynamic Light Scattering (DLS)

Through the measurement of the diffusion coefficient, dynamic light scattering also gives access to the hydrodynamic radius of proteins, which depends upon their size and their shape. Globular proteins differ notably from fully or partly unstructured proteins in their hydrodynamic properties, and an empirical relationship between number of residues N and hydrodynamic radius R_s has been established for native, globular proteins and for fully denatured proteins (Wilkins et al., 1999). The hydrodynamic radius measured for PNT by light scattering analysis in 100 mM NaCl at pH 8 at a concentration of 1 mg/ml was $R_s = 47 \pm 4$ Å (the same value was found in the presence of 10 mM DTT) which is consistent (within the error bars) with the R_s measured by gel filtration. According to (Wilkins et al., 1999), the theoretical radius expected for a native, globular protein with 236 aa as measured by DLS or by pulse field gradient NMR techniques is about 25 Å, whereas for a fully denatured protein, it would be close to 50 Å. Thus, the R_s of PNT measured by DLS corresponds to the value expected for a fully unfolded protein.

Folding of PNT induced by TFE

The regions of PNT involved in association with other proteins may undergo disorder-to-order transitions upon binding to their target, a process termed «induced folding». To identify such potential regions, the solvent trifluoroethanol (TFE) was used. TFE is widely used as an empirical probe of hidden structural propensities of peptides and proteins (Hua et al., 1998). In particular, activators of transcription, whose similarity with PNT has been mentioned above, are involved in protein-protein interactions and might therefore not experience an

aqueous environment *in vivo*. Therefore, in a study of the transcriptional activation domain of Thyroid Transcription Factor-1 TTF-1, TFE was used to mimic the hydrophobic milieu they might be experiencing (Tell et al., 1998). Indeed, TFE induced the formation of helical structure in a region of TTF-1 known for its biological function. Other studies have also validated the use of TFE to uncover regions that have a propensity to undergo an induced folding (Dahlman-Wright and McEwan, 1996; Hua et al., 1998).

CD spectra of PNT were recorded in the presence of concentrations of TFE ranging from 0 to 50% (Figure 6A). PNT shows an increasing gain of α-helicity upon the addition of TFE, as evidenced by the characteristic maximum at 190 nm and minima at 208 and 222 nm. Most unstructured/structured transitions take place before 20% TFE, a concentration at which the α-helix content is estimated to be about 15% (using the ellipticity at 220 nm). This behavior has been reported for the VP16 AAD (Donaldson and Capone, 1992), but is not general, since for instance the GCN4 AAD forms little or no α-helix in TFE concentrations as high as 30%, and folds in fact mostly as β-sheets in 50% TFE at pH 5.5 (Van Hoy et al., 1993). Thus, the behavior of PNT in TFE reveals an α-helix forming potential.

Identification of regions of PNT that fold in TFE

Limited proteolysis is widely used as a probe of protein structure. Globular proteins are preferentially cleaved at exposed and flexible loops only and almost never within secondary structure elements (Fontana et al., 1997; Hubbard, 1998). The use of a protease with broad substrate specificity, such as thermolysin (a TFE-resistant enzyme) allows the identification of cleavage sites solely on the basis of the flexibility of the protein substrate. As a control of the functionality of thermolysin in TFE, cytochrome c type IV was digested by thermolysin in 50% TFE, which yielded two proteolytic fragments of 4 kDa and 7 kDa, as described in (Fontana et al., 1995) (not shown). To map the regions of PNT that gain secondary structure upon addition of TFE, we submitted PNT to digestion by thermolysin in the presence of varying concentrations of TFE. The minimum concentration of TFE necessary to obtain a fragment resistant to proteolysis was found to be 15% (not shown). At this concentration, the α-helix content is estimated at about 10% (between that at 10% TFE and that at 20% TFE), using the ellipticity at 220 nm. A time course experiment in the presence of thermolysin and 15% TFE (Figure 6B) allowed us to identify a digestion-resistant fragment (Figure 6B, arrow 2). In the absence of TFE, PNT is entirely digested in 6 hours, whereas it shows resistance to digestion in the presence of 15 % TFE (compare lane C and 4). The peptidic fragment was isolated and submitted to mass spectrometry and N-terminal sequence analysis. It starts with the sequence LAIEE and has a size of 7498.16 Da, consistent with a 73 aa peptide encompassing the region 27-99 in the sequence of PNT. Noticeably, the N-terminus of this peptide (aa 27-38) is predicted to be an α-helix by the secondary structure program JPRED (Cuff et al., 1998), while no secondary structure element is predicted within the remainder of the region (aa 39-99) (not shown). However since this 73 aa fragment represents about one

third of PNT, the estimate of a 10% α-helix content for PNT means that the fragment cannot be entirely helical (otherwise PNT should have a α-helix content of at least one third, that is 30%).

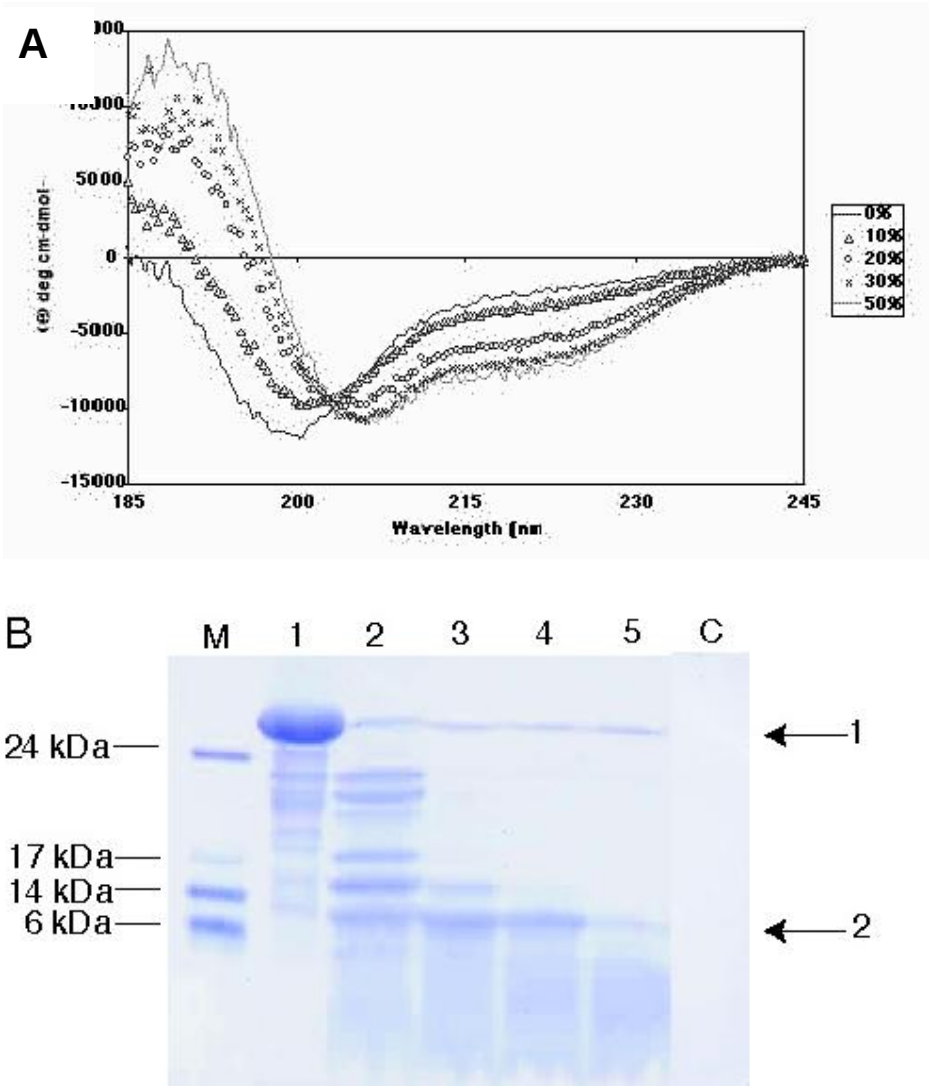

Figure 6. CD spectrum and limited proteolysis of PNT in TFE

A. CD spectrum of PNT in the presence of increasing concentrations of TFE (0, 10, 20, 30 and 50%). B. 15% SDS-PAGE. Limited digestion of PNT by thermolysin in 15% TFE at 25°C at different time intervals (0, 1h, 3h, 6h, and 24h, lanes 1 to 5). Lane C (control): digestion of PNT by thermolysin in 0% TFE, 6h at 25°C. M = Markers. The arrows show undigested PNT (1) and a fragment resistant to proteolysis (2).

Prediction of unstructured regions of *Morbillivirus* PNTs

The term «natively unfolded protein» was coined recently to refer to proteins that have no or little secondary structure under physiological conditions, in the absence of a physiological partner (Weinreb et al., 1996); for a review see (Dunker et al., 2001; Wright and Dyson, 1999). Natively unfolded proteins differ significantly in their primary sequence from small, globular, folded proteins (Uversky, Gillespie, and Fink, 2000). A distinct combination of low hydrophobicity and relatively high net charge allows to predict natively unfolded proteins on the basis of their sequence alone, with very good confidence. This led us to investigate in more detail the physico-chemical properties of PNT in order to check whether it could be

natively unfolded. Indeed, as shown in Table I, PNT seems to fall into the class of natively unfolded proteins, since its mean hydrophobicity (H) and mean net charge (R) obey the equation previously described in (Uversky, Gillespie, and Fink, 2000): (H) < (C) where (C) = ((R)+1.151)/2.785. In the case of PNT, (H) = 0.42, (R) = 0.07, and (C) = 0.44 (Table 1), thus fulfilling the requirement for natively unfolded proteins, namely that (H) < (C). Calculations performed on the equivalent domains of related viruses (Table 1) show that all the N-terminal domains of *Morbillivirus* P proteins are predicted to be natively unfolded. The W protein of SeV, which corresponds to the PNT domain, is also predicted to be unfolded (Table 1). The Domain I of Vesicular Stomatitis Virus (VSV), which might share some functional similarities with PNT (discussed below), was also found to be natively unfolded (Table 1). All these domains are acidic with a low hydrophobicity profile, and consist of 200-320 residues. Conversely, shorter, basic N-terminal domains of *Rubulavirus* P proteins are predicted to be folded (not shown).

The group of K. Dunker and Z. Obradovic developed a series of predictors for the identification of regions within proteins that lack fixed tertiary structures, essentially being partially or fully unfolded and referred to as "disordered regions" (Li et al., 1999; Romero et al., 1997; Romero et al., 2001). While the approach of Uversky *et al.* is useful for wholly unfolded proteins, the programs of the Dunker/Obradovic group can deal with proteins having both ordered and disordered regions. This group developed neural networks trained to distinguish ordered from disordered regions, the disorder of which was assessed using either X-ray crystallography or NMR. This collection of Predictors of Natural Disordered Regions is termed PONDR. The reliability of the prediction increases with the length of the region concerned. A borderline length of 40 residues has been chosen by Dunker *et al.* in their extensive study to define Long Disordered Regions (LDRs) of proteins with very good confidence (Romero et al., 1998). The error rate for prediction of LDRs consisting of 40 aa is 0.4% and falls below distinguishable levels for LDRs longer than 60 aa (Romero et al., 2001). The neural network predictor VL-XT from the program PONDR was run on both MV PNT and on SeV protein W, the functional counterpart of PNT. As shown in Figure 7, two LDRs are predicted within MV PNT, ranging from aa 20-103 and 125-170, with respective strengths of 0.79 and 0.85. Similarly, two LDRs are predicted within SeV W: a first region, ranging from aa 56 to 106 and a second, much longer one (aa 126 to 317) (Figure 7). As mentioned above, the strength of the prediction increases with the length of the LDR and the latter is an extremely strong prediction, with a strength of 0.9 and a length of 191 residues. Comparatively, VL-XT predicts no region of disorder longer than 35 aa in the MV L protein which is 2183 residues long (not shown). In all, more than 240 residues out of 320 are predicted to belong to a disordered region for W, and 130 out of 230 for PNT, thus reinforcing the prediction that was made on the basis of their hydrophobicity and mean net charge.

Taken together, our results suggest that MV PNT is not structured in solution, and that the PNT domains of *Morbilliviruses* and *Paramyxoviruses* are natively unfolded. Folding of PNT can be induced under hydrophobic conditions, and one likely structural element involved in this induced folding is the region encompassing residues 27 to 99, which contains a computer-predicted alpha-helix (residues 27 to 38).

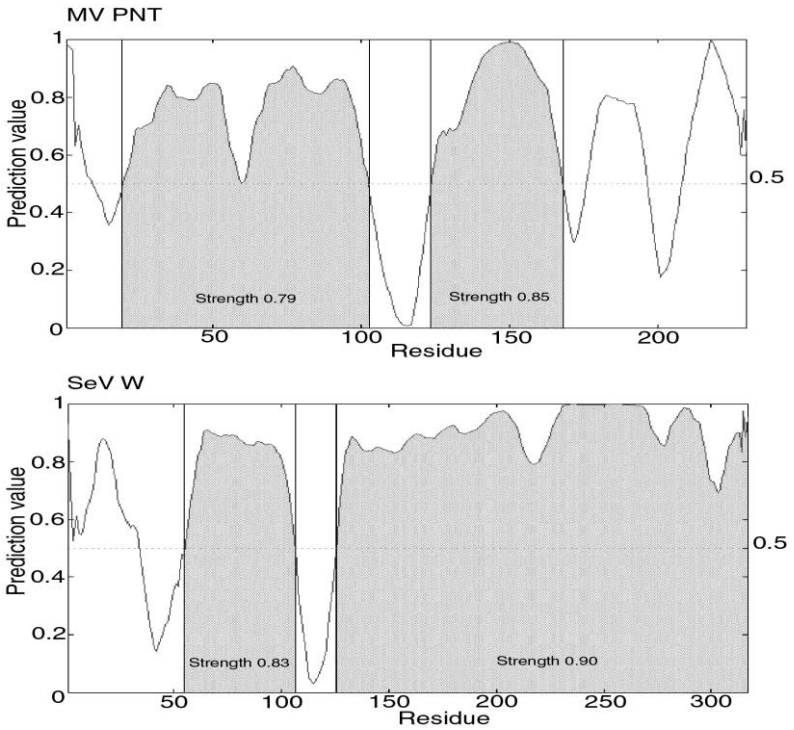

Figure 7. PONDR disorder prediction of MV PNT and of SeV W protein

Disorder prediction values for a given residue are plotted against the residue number. The significance threshold, above which residues are considered to be disordered, set to 0.5, is shown. Long (> 40 residues) Disordered Regions (LDRs) are shaded (residues 20-103 and 125-170 for MV PNT; residues 56-106 and 126-317 for SeV W). The strength of the prediction for each LDR is shown within the corresponding shaded region.

Family	Genus	Species	Size (aa)	pI	Mean Net Charge (R)	Mean Hydrophobicity (H)	(C) = ((R)+1.151)/2.785	Natively unfolded ? (if C >H)
RHABDOVIRIDAE	**Vesiculovirus**	VSV	198	4.1	0.13	0.42	0.45	Yes
PARAMYXOVIRIDAE	**Morbillivirus**	MV	230	4.9	0.07	0.42	0.44	Yes
		CDV	230	4.5	0.02	0.40	0.45	Yes
		RPV	230	4.4	0.12	0.42	0.46	Yes
	Paramyxovirus	SeV	317	5.2	0.04	0.38	0.43	Yes

Table 1. Net charge/hydrophobicity analysis of the PNT domains of *Paramyxoviruses* and *Morbilliviruses*.

All domains considered start at the first residue of the corresponding phosphoprotein and their length is indicated in the fourth column. Mean Net Charge (R) and Mean Hydrophobicity (H) are calculated as described in Material and Methods. A protein is predicted as natively unfolded if (H)<C where (C) =((R)+1.151)/2.785). Abbreviations are as follow: Vesicular Stomatitis virus (VSV) - Measles Virus (MV) - Canine Distemper Virus (CDV) - Rinderpest Virus (RPV) - Sendai Virus (SeV). Accession numbers are given under Material and Methods.

DISCUSSION

We report the expression and purification of the N-terminal domain of the phosphoprotein of measles virus, PNT. PNT is found in at least 2 proteins in MV, P and V, and in a third one, W, in the case of SeV. Contrary to the SeV W protein, MV PNT is never expressed alone but always as the N-terminal domain of either P or V. As PNT is soluble and stable, it is reasonable to assume that it constitutes a domain on its own.

The hydrodynamic properties of PNT inferred from both gel filtration and dynamic light scattering experiments are consistent with PNT being either a stable oligomer, or a monomeric, unfolded protein. NMR data show that PNT exhibits no short-range interactions, indicating the lack of any constant secondary structure as well as the absence of any stable oligomers. Moreover, CD studies further confirm that PNT is unfolded in aqueous solution. Taken together, all these experiments show that PNT is unstructured under our experimental conditions, mimicking physiological conditions in terms of pH and salinity.

A recent study showed that it is possible to discriminate with very good confidence between globular proteins and natively unfolded proteins on the basis of their sequence alone (Uversky, Gillespie, and Fink, 2000). Drawing on these results, we show that MV PNT has a combination of mean hydrophobicity and net charge that is typical of the class of natively unfolded proteins. We extend this result to the PNTs of other *Morbilliviruses* and to the W protein of SeV, which is the functional counterpart of PNT in *Paramyxoviruses*. We also use another approach to predict the location of intrinsically disordered regions of proteins, based on the program PONDR (Li et al., 1999; Romero et al., 2001). PONDR identified 2 LDRs in MV PNT, which together span more than half the domain. SeV W also contains 2 LDRs, one of them strikingly long, extending by itself for more than half the length of the protein; in all, more than two thirds of W are predicted to be disordered. These results reinforce and explicit for both domains the predictions that were made on the basis of their mean hydrophobicity and net charge. Furthermore, our results highlight the complementarity of the two methods. Indeed, Uversky *et al.* (Uversky, Gillespie, and Fink, 2000) report that the 19 strongest predictions of LDRs described in (Romero et al., 1998) all possess the characteristics of natively unfolded proteins.

The N-terminal domains of *Rubulavirus* P and V proteins are predicted to be folded. This is in agreement with the idea that they are biochemically and functionally distinct from the *Morbillivirus* and *Paramyxovirus* PNT domains (Lamb and Kolakofsky, 2001).

Sequence variability has been suggested as another possible indicator of structural disorder (Dunker et al., 1998). Contrary to the previous approaches, based on the analysis of a single sequence, it depends upon the analysis of a set of related sequences. In this frameship, we note that well-conserved regions of P (located within PCT) are structured (Tarbouriech et al., 2000), while we show here that the hypervariable N-terminal domain of P, PNT, is

disordered. Thus, our results support the idea that sequence variability might be an indicator of structural disorder.

The peculiar sequence properties of N-terminal domains of phosphoproteins have long been noticed as reminiscent of AADs from cellular transcriptional factors. The structure and function of AADs is still unclear but AADs play a role in recruiting the transcriptional machinery via protein-protein interactions (Melcher, 2000). Intriguingly, AADs can be functionally replaced by a number of unrelated proteins, they act in a position-independent manner, and their function does not depend upon a distinct tertiary structure. Recently, several studies have shown that AADs act through bulky hydrophobic residues scattered within acidic residues (Melcher, 2000). The numerous, charged residues would help to keep these hydrophobic residues in an aqueous environment, allowing them to establish weak, short distance contacts with hydrophobic patches in their targets. In that model, i) specificity of AADs for their natural partners is provided by co-localization with their targets (on DNA); ii) affinity is attained through to the use of multiple activation domains. This occurs through either multimerization of transcriptional factors containing AADs, or by combining within an AAD several redundant, independent regions of interaction, each of low affinity (Melcher, 2000).

The first notice of the similarity between a phosphoprotein N-terminal domain and AADs came from studies of the N-terminal domain (domain I) of the related *Rhabdoviridae* VSV P protein, which is required for mRNA synthesis *in vitro*. This domain can be functionally replaced with the sequence of tubulin, and the 44 aa carboxy terminus of tubulin is indispensable for this function (Chattopadhyay and Banerjee, 1988). Interestingly, we note that this latter region is made of acidic amino acids interspersed with bulky hydrophobic residues, similarly to AADs, providing a possible explanation for its role in substituting P domain I. In the same vein, two regions of SeV PNT essential for genomic RNA synthesis act in a position-independent manner and can partially substitute each other (Curran, Pelet, and Kolakofsky, 1994). Such amino-terminal regions have also been identified in MV and RPV PNT (Curran, Marq, and Kolakofsky, 1995; Harty and Palese, 1995; Shaji and Shaila, 1999). On the basis of its primary sequence, we predict that the VSV P domain I is natively unfolded (Table I), thus reinforcing the proposed structural similarities between VSV P domain I, *Paramyxoviridae* PNTs, and AADs.

Most, though not all, natively unfolded proteins or protein domains characterized so far undergo some degree of folding in the presence of a physiological partner, a process termed «induced folding» (for a review see Dunker et al., 2001; Wright and Dyson, 1999). So far two proteins that bind to PNT have been identified: the unassembled form (N°) of the nucleoprotein and the polymerase (L). N° binds to PNT both within P and V (Curran, Marq, and Kolakofsky, 1995; Harty and Palese, 1995; Shaji and Shaila, 1999; Tober et al., 1998). However, it has not been possible to test the folding of PNT induced by N° using

recombinant PNT and N expressed separately, because in heterologous expression systems N is only produced in its self-assembled form N^{NUC} (Warnes et al., 1995; Warnes, Fooks, and Stephenson, 1994). In addition, V binds to L in the case of the *Morbillivirus* RPV (Sweetman, Miskin, and Baron, 2001), and expression of recombinant L would be necessary to further examine this interaction.

The hidden structural propensities of unstructured proteins can be probed with the solvent TFE, which mimics hydrophobic conditions that such proteins might experience when binding hydrophobic patches in their targets. The specificity of TFE as an α-helix inducing solvent has been validated in a number of studies, which have often established a link between α-helical structure gain upon addition of TFE and functional activity (Dahlman-Wright and McEwan, 1996; Hua et al., 1998; Tell et al., 1998). Using TFE, we show that PNT has a clear α-helix forming potential. We identified one fragment, which spans aa 27-99, resistant to proteolysis in 15% TFE. Noticeably, it belongs to the region 20-103 of PNT, predicted as being disordered by PONDR. Remarkably, the N-terminus of this fragment is predicted to be an α-helix (aa 27-38) by the program JPRED (Cuff et al., 1998). It is reasonable to assume that under these conditions, it is folded as an α-helix, thus preventing cleavage and digestion by thermolysin. Thus, it may represent one of the secondary structure elements involved in the unstructured/structured transition of PNT upon binding to a physiological partner.

The distribution of the conformations of natively unfolded PNT is narrow, according to the relative sharpness of the elution peak observed in gel filtration, compared to what would be expected for a fully denatured protein in the presence of chaotropic agents (urea, guanidinium chloride). This result suggests that some residual structure may prevent the polypeptide chain from adopting multiple conformations. One can then reasonably assume that these residual interactions enable a more efficient start of the folding process induced by a binding partner.

So far the only detailed mapping of PNT function in replication comes from studies conducted on SeV. Two functions have been uncovered in SeV PNT: a) a chaperone activity for N (mapped to aa 33-41), involving binding to N° and preventing its self-assembly; b) a function in genomic RNA synthesis, whose precise nature is unknown, via two redundant regions reminiscent of AADs (aa 1-77 and 78-156) (Curran, Marq, and Kolakofsky, 1995; Curran, Pelet, and Kolakofsky, 1994). It is tempting to relate the latter function to the L-PNT interaction recently proven in the case of RPV V (Sweetman, Miskin, and Baron, 2001). Indeed, several evidences point out that SeV PNT interacts with L (Curran, Marq, and Kolakofsky, 1995; Curran, Pelet, and Kolakofsky, 1994; Sedlmeier and Neubert, 1998). Although attempts to demonstrate a direct PNT-L interaction for SeV by immunoprecipitation failed so far, we note that AADs generally have a weak affinity for their targets (Melcher, 2000). Given the similarity of PNT to AADs, such an interaction might have escaped

detection. On the other hand, a stable binding site for L has been mapped to the PCT domain (Smallwood, Ryan, and Moyer, 1994). Thus we note a parallel with two important features of the mode of action of AADs mentioned above: i) co-localization of PNT with L on the (N:RNA) complex is provided by an independent and stable L-binding site on P. ii) the use of multiple activation domains is provided either by multimerization of P and/or by the presence of redundant activation regions in PNT.

In addition to viral proteins, MV V binds three different, uncharacterized cellular proteins in infected cells, with both the N-terminal (PNT) and the C-terminal (zinc binding) domains of V being required for this interaction (Liston, DiFlumeri, and Briedis, 1995). Thus PNT plays very different roles during viral replication. Indeed, one advantage of unstructured regions of proteins might be their ability to bind to a wide variety of structurally-distinct substrates (Dunker et al., 2001; Wright and Dyson, 1999), as already proposed for SeV PNT (Curran, Pelet, and Kolakofsky, 1994). Although speculative, it is possible that one role of PNT is to put into contact several proteins within the replicative assembly complex, such as N° and L.

In this study, we have shown using biochemical and biophysical methods that the recombinant PNT is not structured. Its sequence properties, characteristic of natively unfolded proteins, suggest that this is an intrinsic property of PNT. However, further studies, involving for instance co-expression of P with other viral proteins, are necessary to establish the biological relevance of these observations.

MATERIAL AND METHODS

Bacterial strains and media

The *E. coli* strain XL1 Blue (Stratagene, CA) was used for selection and amplification of DNA constructs. The *E. coli* strain BL21(DE3) (Novagen) was used for expression of recombinant proteins. *E. coli* was grown in Luria-Bertani medium.

Chemicals

Restriction enzymes and T4 DNA ligase were purchased from NE Biolabs (USA). *Pfu* polymerase was from Promega (USA). Primers were purchased from Genosys (UK). Nucleotide Sequencing was carried out by Genome Express (France).

Cloning of the P and PNT coding regions

The plasmid pSC6/P encodes the measles virus P gene (strain Edmonston B) (Radecke et al., 1995). P was cloned into the expression vector pET21a (Novagen) by PCR using pSC6/P as a template. Forward primer (5' GGCCCATATGGCAGAAGAGCAGGCACGC 3') was designed to introduce an *NdeI* restriction site fused to the N-terminus of P, and reverse primer (5' CCGGGAATTCTTACTTCATTATTATCTTCAT 3') was designed to introduce an *EcoRI* site after the P stop codon. The PCR product was digested with *NdeI* and *EcoRI*, purified, and ligated between the *NdeI* and *EcoRI* sites of pET21a to yield pET21a/P. To introduce an hexahistidine tag fused to the N-terminus of P, pET21a/P was digested with *Nde*I and *BamH*I, dephosphorylated, and ligated to a double-stranded, phosphorylated synthetic oligonucleotide (5' TATGCCCATCATCATCATCATGCAGAAGAGCAGGCACGCCATG TCAAAAACGGACTCGAGTGCA 3' annealed to 5' CTCGAGTCCGTTTTTGACTGGCG TGCCTGCTCTTCTGCATGATGATGATGATGGTGCA-3') to yield pET21a/P_{H6}.

The PNT gene construct encoding residues 1-230 of P with an hexahistidine tag fused to its N-terminus was obtained by inserting a stop codon in position 691 of the sequence of the P ORF by PCR, using pET21/P_{H6} as template, forward primer (5' GGCCAGGCCTTGATGGTGATAGCAC 3') and reverse primer (5' GGCCGGATCCGGTACCTTATTACTTTTTAATGGGTGTCCC GGAAGT 3'). The PCR product was digested with *StuI* and *BamHI*, purified, and ligated between the *StuI* and *BamHI* sites of the pET21a/P_{H6} plasmid to yield pET21a/PNT_{H6}.The sequence of all coding regions was verified by nucleotide sequencing.

Expression of PNT

E. coli BL21(DE3) transformed with pET21a/PNT_{H6} and with pUBS520 (Brinkmann, Mattes, and Buckel, 1989) was grown overnight to saturation in Luria-Bertani (LB) medium containing 100µg/mL ampicilin and 50µg/mL kanamycin. An aliquot of the overnight culture was diluted 1/25 in LB medium and grown at 37°C. At OD_{600} of 0.7, isopropyl ß-D-thiogalactopyranoside

(IPTG) was added to a final concentration of 0.2 mM, and the cells were grown at 37°C for 3 hours. The induced cells were harvested, washed and collected by centrifugation. The resulting pellet was frozen at –20°C.

Purification of PNT

The pellet containing PNT was resuspended in 5 volumes (w/w) buffer A (50 mM Sodium Phosphate pH 8, 300 mM NaCl, 1 mM PMSF) supplemented with 1 mM benzamidine, 10 mM imidazole, 100 µg/ml lysozyme, 1 µg/ml DNAse I. After a 20 min incubation on ice, the cells were disrupted by sonication (using a 750 W sonicator and 4 cycles of 50 s at 50% power output). The lysate was clarified by centrifugation at 30 000 g for 30 min at 4°C.

Starting from a 1 l culture, the clarified supernatant was incubated for 1 hr at 4°C with gentle shaking with 4 ml Chelating Sepharose Fast Flow Resin preloaded with Ni^{2+} ions (Amersham Pharmacia Biotech), previously equilibrated in buffer A containing 10 mM imidazole. The resin was washed with buffer A containing 20 mM imidazole, and PNT was eluted in buffer A containing 250 mM imidazole. Eluates were analyzed by SDS-PAGE for the presence of PNT. The fractions containing PNT were combined, diluted with 5 volumes of 10 mM Tris pH 8, 1 mM PMSF and loaded onto a Source 15Q PE 4.6/100 column (Amersham Pharmacia Biotech) equilibrated in buffer B (10 mM Tris pH 8, 50 mM NaCl, 1 mM PMSF). Elution was carried out with a gradient of NaCl (50 mM-500 mM) in buffer B. Fractions containing PNT were pooled and loaded onto a Superdex 200 HR 10/30 column (Amersham Pharmacia Biotech) and eluted with buffer C (10 mM Tris pH 8, 50 mM NaCl). The protein was concentrated using Centripreps 10 (Millipore). Apparent molecular weights of proteins eluted from the column were deduced from a calibration carried out with LMW and HMW calibration kits (Amersham Pharmacia Biotech). The theoretical Stokes radii of a native (R_sN) and fully unfolded (R_sU) protein with a molecular weight MW (in Daltons) were calculated according to (Uversky, 1993): $\log(R_sN) = 0.369*\log(MW) - 0.254$ and $\log(R_sU) = 0.533*\log(MW) - 0.682$.

Circular Dichroism

The CD spectra were recorded on a Jasco 800 dichrograph using 1-mm thick quartz cells in 10 mM Sodium Phosphate buffer pH 7.0 at 20°C. Structural variations were measured as a function of changes in the initial CD spectrum upon addition of increasing concentrations of 2,2,2-trifluoroethanol (TFE) (Fluka). CD spectra were measured between 185 and 260 nm. Mean ellipticity values per residue ([Θ]) were calculated as $[\Theta] = 3300\ m\ \Delta A / (l\ c\ n)$, taking into account a path length (l), a number of residues (n), a molecular mass (m), and a protein concentration (c) of respectively 0.1 cm, 236 25 kDa and 0.2 mg/ml. The alpha-helix content was derived from the ellipticity values at 220 nm as described in (Morris et al., 1999).

Proteolysis with thermolysin

Horse heart cytochrome c (type VI) and thermolysin were obtained from Sigma. Digestion of proteins was performed by incubation of the protein substrate (1mg/ml) dissolved in 20 mM Tris/HCl pH 7.8, with thermolysin (protease/protein substrate ratio 1: 100) at 25°C. The

reaction was carried out in the absence or presence of TFE (from 10 up to 50%). The extent of proteolysis was evaluated by SDS-PAGE analysis on 20 µl aliquots that were removed from the reaction mixture, added to 20 µl of 2 x Laemmli sample buffer and boiled for 5 min to inactivate the protease.

Dynamic Light scattering (DLS)

Dynamic light scattering experiments were performed with a Dynapro MSTC-200 (Protein Solutions) at 20°C. All samples were filtered prior to the measurements (Whatman Anodisc 13, 0.02 μm). For total intensity light scattering, protein concentrations were determined by measuring absorbance at 280 nm directly in the light scattering sample cell and the instrument was calibrated with toluene using the same cell.

Hydrodynamic radii were deduced from translational diffusion coefficients using the Stokes-Einstein equation. Diffusion coefficients were inferred from the analysis of the decay of the scattered intensity autocorrelation function for monodisperse samples. All calculations were performed using the software provided by the manufacturer.

Mass Spectrometry

Mass analysis of PNT was performed using a Voyager DE RP mass spectrometer (PerSeptive Biosystems). Samples (0.7µl containing 15 pmol) were mixed with an equal volume of sinapinic acid matrix solution, spotted on the target, then dried at room temperature for 10 min. The mass standard was apomyoglobin.

Two-dimensional NMR

A sample containing purified PNT at a concentration of 0.3 mM in 50 mM Sodium Acetate pH 5.6, 30 mM NaCl, was used for the acquisition of a NOESY spectrum on a DRX500 Bruker spectrometer at 300 K with 2048 complex points in the directly acquired dimension and 512 complex points in the indirectly detected dimension. Solvent suppression was achieved by the WATERGATE 3-9-19 pulse (Piotto, Saudek, and Sklenar, 1992). The data were processed using the UXNMR software; they were multiplied by a sine-squared bell and zero-filled to 1 K in ω1 dimension prior to Fourier transformation.

Sequence Retrieval

Sequences for this study were obtained using the National Center for Biotechnology Information's (NCBI) sequence retrieval software. *Paramyxoviridae* and *Rhabdoviridae* P sequence accession numbers are as follows:

Vesicular Stomatitis virus (VSV) - AAA48451; Measles Virus (MV) - CAA91364; Canine Distemper Virus (CDV) - AAG15488; Rinderpest Virus (RPV) - JQ1930; Sendai Virus (SeV) - P04860.

Calculation of Mean Net Charge (R) and Mean Hydrophobicity (H)

The Mean Net Charge of a protein is determined as the absolute value of the difference between the number of positively and negatively charged residues divided by the total

number of amino acid residues. It was calculated using the program ProtParam at the EXPASY server (http://www.expasy.ch/tools).

The Mean Hydrophobicity (H) is the sum of normalized hydrophobicities of individual residues divided by the total number of amino acid residues minus 4 residues (to take into account fringe effects in the calculation of hydrophobicity). Individual hydrophobicities were determined using the Protscale program at the EXPASY server (http://www.expasy.ch/tools), using the options "Hphob / Kyte & Doolittle", a window size of 5, and normalizing the scale from 0 to 1. The values computed for individual residues were then exported to a spreadsheet, summed and divided by the total number of residues minus four to yield (H).

PONDR prediction of unstructured regions

Sequences of proteins were submitted to the PONDR server (http://www.disorder.wsu.edu/) using the default integrated predictor VL-XT (Li et al., 1999; Romero et al., 2001). Access to PONDR was provided by Molecular Kinetics (P.O. Box 2475 CS, Pullman, WA 99165-2475; E-mail: mkinetics@turbonet.com) under license from the WSU Research Foundation.

ACKNOWLEDGEMENTS

We thank H. Van Tilbeurgh for timely advice, H. Darbon for help with the NMR study, F. Heitz for help with circular dichroism, F. Hoh and S. Bressanelli for help with DLS, as well as V. Uversky and K. Dunker for useful comments and critical reading of the manuscript; this work was supported by the CNRS (ATIPE).

REFERENCES

Brinkmann, U., Mattes, R. E., and Buckel, P. (1989). High-level expression of recombinant genes in Escherichia coli is dependent on the availability of the dnaY gene product. *Gene* 85, 109-14.

Cattaneo, R., Kaelin, K., Baczko, K., and Billeter, M. A. (1989). Measles virus editing provides an additional cysteine-rich protein. *Cell* 56, 759-64.

Chattopadhyay, D., and Banerjee, A. K. (1988). NH2-terminal acidic region of the phosphoprotein of vesicular stomatitis virus can be functionally replaced by tubulin. *Proc. Natl. Acad. Sci.* U S A 85, 7977-81.

Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., and Barton, G. J. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics* 14, 892-3.

Curran, J. (1998). A role for the Sendai virus P protein trimer in RNA synthesis. *J. Virol.* 72, 4274-80.

Curran, J., Boeck, R., Lin-Marq, N., Lupas, A., and Kolakofsky, D. (1995). Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils. *Virology* 214, 139-49.

Curran, J., and Kolakofsky, D. (1999). Replication of paramyxoviruses. *Adv. Virus Res.* 54, 403-22.

Curran, J., Marq, J. B., and Kolakofsky, D. (1995). An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J. Virol.* 69, 849-55.

Curran, J., Pelet, T., and Kolakofsky, D. (1994). An acidic activation-like domain of the Sendai virus P protein is required for RNA synthesis and encapsidation. *Virology* 202, 875-84.

Dahlman-Wright, K., and McEwan, I. J. (1996). Structural studies of mutant glucocorticoid receptor transactivation domains establish a link between transactivation activity in vivo and alpha-helix-forming potential in vitro. *Biochemistry* 35, 1323-7.

Delenda, C., Hausmann, S., Garcin, D., and Kolakofsky, D. (1997). Normal cellular replication of Sendai virus without the trans-frame, nonstructural V protein. *Virology* 228, 55-62.

Delenda, C., Taylor, G., Hausmann, S., Garcin, D., and Kolakofsky, D. (1998). Sendai viruses with altered P, V, and W protein expression. *Virology* 242, 327-37.

Donaldson, L., and Capone, J. P. (1992). Purification and characterization of the carboxyl-terminal transactivation domain of Vmw65 from herpes simplex virus type 1. *J. Biol. Chem.* 267, 1411-4.

Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., and Villafranca, J. E. (1998). Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac. Symp. Biocomput.,* 473-84.

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001). Intrinsically disordered protein. *J. Mol. Graph. Model* 19, 26-59.

Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1997). Probing the partly folded states of proteins by limited proteolysis. *Fold. Des.* 2, R17-26.

Fontana, A., Zambonin, M., De Filippis, V., Bosco, M., and Polverino de Laureto, P. (1995). Limited proteolysis of cytochrome c in trifluoroethanol. *FEBS Lett.* 362(3), 266-70.

Gombart, A. F., Hirano, A., and Wong, T. C. (1993). Conformational maturation of measles virus nucleocapsid protein. *J. Virol.* 67, 4133-41.

Gualfetti, P. J., Iwakura, M., Lee, J. C., Kihara, H., Bilsel, O., Zitzewitz, J. A., and Matthews, C. R. (1999). Apparent radii of the native, stable intermediates and unfolded conformers of the alpha-subunit of tryptophan synthase from E. coli, a TIM barrel protein. *Biochemistry* 38, 13367-78.

Harty, R. N., and Palese, P. (1995). Measles virus phosphoprotein (P) requires the NH2- and COOH-terminal domains for interactions with the nucleoprotein (N) but only the COOH terminus for interactions with itself. *J. Gen. Virol.* 76, 2863-7.

Horikami, S. M., Smallwood, S., and Moyer, S. A. (1996). The Sendai virus V protein interacts with the NP protein to regulate viral genome RNA replication. *Virology* 222, 383-90.

Hua, Q. X., Jia, W. H., Bullock, B. P., Habener, J. F., and Weiss, M. A. (1998). Transcriptional activator-coactivator recognition: nascent folding of a kinase-inducible transactivation domain predicts its structure on coactivator binding. *Biochemistry* 37, 5858-66.

Hubbard, S. J. (1998). The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta* 1382, 191-206.

Kato, A., Kiyotani, K., Sakai, Y., Yoshida, T., Shioda, T., and Nagai, Y. (1997). Importance of the cysteine-rich carboxyl-terminal half of V protein for Sendai virus pathogenesis. *J. Virol.* 71, 7266-72.

Lamb, R. A., and Kolakofsky, D. (2001). Paramyxoviridae: The Viruses and Their Replication. *In "*Fields Virology*"* (B.N. Fields, D.M. Knipe, and P.M. Howley, Eds.), 4th ed., pp 1305-1340. Lippincott-Raven, Philadelphia, PA.

Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999). Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform. Ser. Workshop Genome Inform.* 10, 30-40.

Liston, P., and Briedis, D. J. (1994). Measles virus V protein binds zinc. *Virology* 198, 399-404.

Liston, P., DiFlumeri, C., and Briedis, D. J. (1995). Protein interactions entered into by the measles virus P, V, and C proteins. *Virus Res.* 38, 241-59.

Melcher, K. (2000). The strength of acidic activation domains correlates with their affinity for both transcriptional and non-transcriptional proteins. *J. Mol. Biol.* 301, 1097-112.

Morris, M. C., Mery, J., Heitz, A., Heitz, F., and Divita, G. (1999). Design and synthesis of a peptide derived from positions 195-244 of human cdc25C phosphatase. *J. Pept. Sci.* 5, 263-71.

Patterson, J. B., Thomas, D., Lewicki, H., Billeter, M. A., and Oldstone, M. B. (2000). V and C proteins of measles virus function as virulence factors in vivo. *Virology* 267, 80-9.

Piotto, M., Saudek, V., and Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* 2, 661-5.

Precious, B., Young, D. F., Bermingham, A., Fearns, R., Ryan, M., and Randall, R. E. (1995). Inducible expression of the P, V, and NP genes of the paramyxovirus simian virus 5 in cell lines and an examination of NP-P and NP-V interactions. *J. Virol.* 69, 8001-10.

Radecke, F., Spielhofer, P., Schneider, H., Kaelin, K., Huber, M., Dotsch, C., Christiansen, G., and Billeter, M. A. (1995). Rescue of measles viruses from cloned DNA. *EMBO J.* 14, 5773-84.

Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., and Dunker, A. K. (1997). Identifying disordered regions in proteins from amino acid sequences. *Proc. I.E.E.E. International Conference on Neural Networks*, 90-95.

Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., and Dunker, A. K. (1998). Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.,* 437-48.

Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001). Sequence complexity of disordered proteins. *Proteins* 42, 38-48.

Ryan, K. W., Morgan, E. M., and Portner, A. (1991). Two noncontiguous regions of Sendai virus P protein combine to form a single nucleocapsid binding domain. *Virology* 180, 126-34.

Schneider, H., Kaelin, K., and Billeter, M. A. (1997). Recombinant measles viruses defective for RNA editing and V protein synthesis are viable in cultured cells. *Virology* 227, 314-22.

Sedlmeier, R., and Neubert, W. J. (1998). The replicative complex of paramyxoviruses: structure and function. *Adv*. *Virus Res.* 50, 101-39.

Shaji, D., and Shaila, M. S. (1999). Domains of Rinderpest virus phosphoprotein involved in interaction with itself and the nucleocapsid protein. *Virology* 258, 415-24.

Smallwood, S., Ryan, K. W., and Moyer, S. A. (1994). Deletion analysis defines a carboxyl-proximal region of Sendai virus P protein that binds to the polymerase L protein. *Virology* 202, 154-63.

Sweetman, D. A., Miskin, J., and Baron, M. D. (2001). Rinderpest virus C and V proteins interact with the major (L) component of the viral polymerase. *Virology* 281, 193-204.

Tarbouriech, N., Curran, J., Ruigrok, R. W., and Burmeister, W. P. (2000). Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nat. Struct. Biol.* 7, 777-81.

Tell, G., Perrone, L., Fabbro, D., Pellizzari, L., Pucillo, C., De Felice, M., Acquaviva, R., Formisano, S., and Damante, G. (1998). Structural and functional properties of the N transcriptional activation domain of thyroid transcription factor-1: similarities with the acidic activation domains. *Biochem. J.* 329, 395-403.

Tober, C., Seufert, M., Schneider, H., Billeter, M. A., Johnston, I. C., Niewiesk, S., ter Meulen, V., and Schneider-Schaulies, S. (1998). Expression of measles virus V protein is associated with pathogenicity and control of viral RNA synthesis. *J. Virol.* 72, 8124-32.

Uversky, V. N. (1993). Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry* 32, 13288-98.

Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions ? *Proteins* 41, 415-427.

Van Hoy, M., Leuther, K. K., Kodadek, T., and Johnston, S. A. (1993). The acidic activation domains of the GCN4 and GAL4 proteins are not alpha helical but form beta sheets. *Cell* 72, 587-94.

Warnes, A., Fooks, A. R., Dowsett, A. B., Wilkinson, G. W., and Stephenson, J. R. (1995). Expression of the measles virus nucleoprotein gene in Escherichia coli and assembly of nucleocapsid-like structures. *Gene* 160, 173-8.

Warnes, A., Fooks, A. R., and Stephenson, J. R. (1994). Production of measles nucleoprotein in different expression systems and its use as a diagnostic reagent. *J. Virol. Methods* 49, 257-68.

Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., and Lansbury, P. T., Jr. (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* 35, 13709-15.

Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A., and Smith, L. J. (1999). Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* 38, 16424-31.

Wright, P. E., and Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321-31.

COMMENTAIRE

L'étude des protéines désordonnées est en pleine émergence. Les outils expérimentaux qui permettent d'étudier le désordre structural sont pour la plupart relativement anciens et se répartissent en plusieurs catégories : biochimique (protéolyse limitée), biophysiques (cristallographie aux rayons X, dichroïsme circulaire, RMN, SAXS, diffusion de la lumière, gel-filtration - les 3 dernières méthodes donnent accès à des paramètres hydrodynamiques des protéines). Pourtant, nous manquons des bases conceptuelles pour utiliser au mieux ces outils puissants afin d'appréhender la nature et la fonction du désordre structural. L'hétérogénéité des outils de caractérisation du désordre est reflétée par les différentes classifications des protéines désordonnées, qui peut être source de confusion.

- Cette classification peut être fondée par exemple sur des caractéristiques hydrodynamiques. Parmi les protéines ne possédant pas de structure secondaire constante, on distingue ainsi les protéines qui ont une **conformation complètement étendue** en solution (ce qui est clairement le cas de PNT) de celles qui conservent une **certaine compacité** (appelées "molten globules", littéralement "globules fondus"). On observe souvent des "molten globules" lorsqu'on dénature partiellement des protéines, par un pH extrême ou un extrême de température, mais il existe aussi des protéines dont c'est l'état naturel dans des conditions physiologiques *in vitro* (Dunker *et coll.*, 2001) (toutefois, notons que l'usage du terme "molten globule" pour ces dernières est discutée). Une autre classification dont les fondements conceptuels sont totalement différents (elle repose sur la présence ou non d'un cœur hydrophobe dans la protéine) conduit aussi en pratique à la distinction entre "molten globules" et protéines complètement étendues (Tsai *et coll.*, 2001).

- On peut aussi classer les protéines désordonnées sur la base de leur capacité à se structurer (Zetina, 2001) en présence d'un partenaire : (acide nucléique ou protéine), ou bien suite à une modification post-traductionnelle telle que glycosylation ou phosphorylation, ou encore suite à un changement de pH ou de température (Uversky *et coll.*, 2000b; Uversky *et coll.*, 1999). Cette dernière classification est manifestement artificielle et dépend de l'état de nos connaissances.

On voit qu'aujourd'hui, la seule classification rigoureuse est celle qui est fondée sur des critères hydrodynamiques. Une classification fondée sur la *flexibilité* des protéines serait aussi intéressante. Sans doute les progrès techniques permettront-ils d'avoir accès à d'autres paramètres physico-chimiques qui amélioreront notre connaissance des protéines désordonnées. Une question qui nous a préoccupé pendant la réalisation des expériences

de cet article est : les différentes techniques spectroscopiques caractérisent-elles le même type de désordre structural ? En d'autres termes, quelle est la signification du désordre observé par une expérience de CD, ou de RMN ? L'étude de Williams *et coll.* suggère que la RMN et le CD donnent des résultats similaires, alors que la cristallographie aux rayons X fournit des résultats légèrement différents (Williams *et coll.*, 2001). Ceci étant posé, comprendre *exactement* ce qu'est une protéine désordonnée occupe en ce moment même de nombreux biochimistes et biophysiciens de par le monde et sort du cadre de la présente thèse !…

Dans l'état actuel des connaissances, peut-on interpréter *quantitativement* et pas uniquement *qualitativement* le rapport (charge nette moyenne / hydrophobicité moyenne) d'une protéine défini par (Uversky *et coll.*, 2000a) ? Tsai *et coll.* proposent que des protéines ayant une charge nette ou une hydrophobicité extrême aient sans doute une conformation étendue, alors que des protéines proches de la limite entre protéines structurées et désordonnées seraient plus probablement des "molten globules" (Tsai *et coll.*, 2001). Le présent article ne confirme pas une telle interprétation puisque la charge nette et l'hydrophobicité moyennes de PNT (qui possède une conformation étendue) sont toutes deux assez proches des valeurs observées pour des protéines globulaires.

Les rôles possibles du désordre structural de PNT sont abordés en détail dans la revue "Structural flexibility of the replicative complex of *Paramyxoviridae*" au sein de cette thèse. J'aimerais apporter trois compléments à l'article :

1) Les études remarquables de Chinchar et Portner (Chinchar and Portner, 1981) (voir la revue) suggéraient, il y a déjà 20 ans, que P de SeV était divisé en une partie sensible à la protéolyse et donc sans doute désordonnée (PNT), et une partie compacte (PCT), suffisante pour la transcription *in vitro*, le coiffage, la méthylation et la polyadénylation des ARNm. Leurs résultats de protéolyse limitée s'accordent bien avec ceux de (Tarbouriech *et coll.*, 2000).

2) L'article contient une erreur, concernant les deux régions de PNT (aa 1-77 et 78-144) qui auraient un rôle redondant dans la synthèse d'ARN. Leur redondance fonctionnelle, couplée au fait qu'elles semblent fonctionner indépendamment de leur position au sein de P, rappelle les domaines d'activations acides (Curran *et coll.*, 1994). En fait, la fonction attribuée à ces deux régions dans l'article initial (Curran *et coll.*, 1994) était due à un artefact, expliqué dans un article ultérieur (Curran, 1996), ce qui nous a échappé. Ainsi, aucune de ces régions n'a de rôle dans la synthèse d'ARN, puisque PCT seul suffit à assurer la transcription (Curran, 1996). Les 2 régions partagent cependant bien une fonction commune, celle de se lier à N°.

Une explication de l'artefact est présentée en figure 7. Que J. Curran veuille bien trouver ici l'expression de notre regret pour cette malencontreuse erreur.

Figure 7

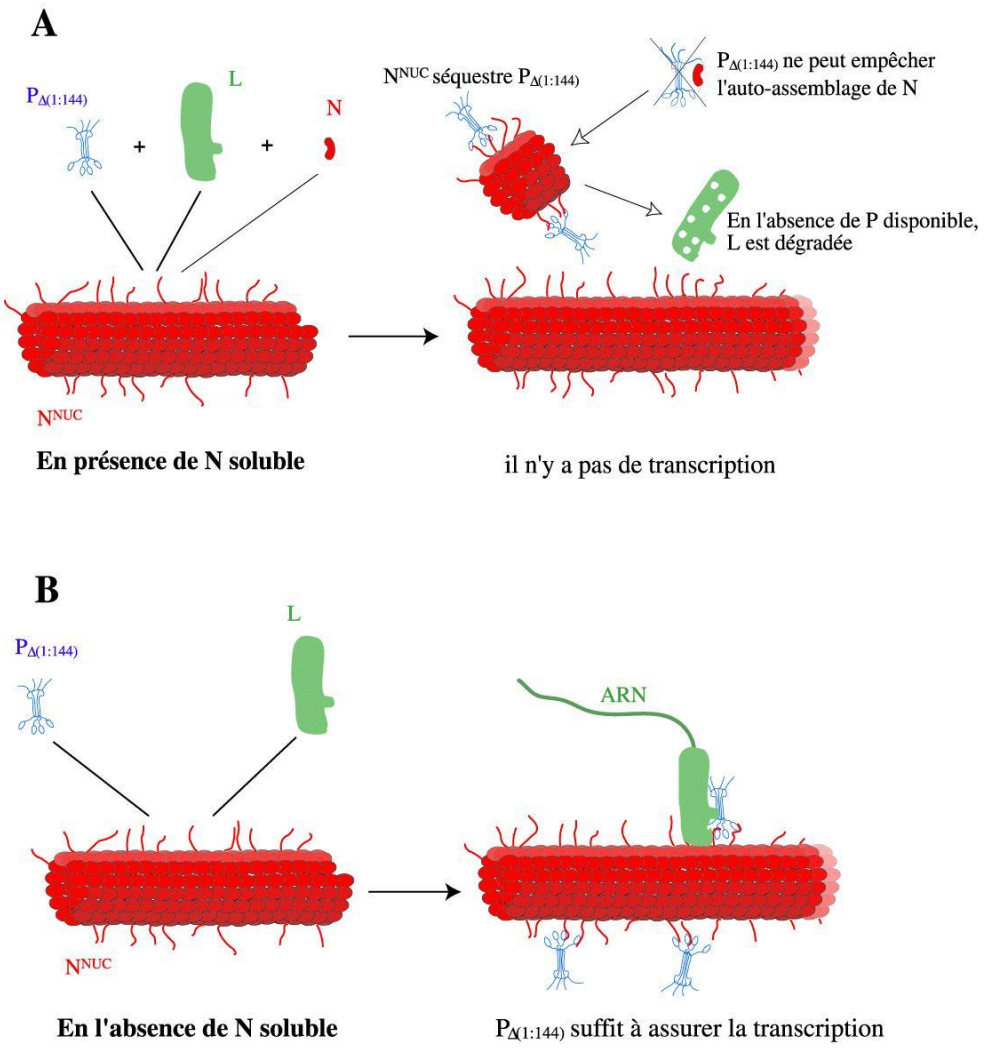

Figure 7. La région 1-144 de P est superflue pour la transcription, sauf en présence de N soluble.

Pour étudier simultanément la transcription et la réplication virale, Curran *et coll.* ont mis au point un système expérimental dans lequel la transcription se déroule en présence de N soluble.

A. En présence de N soluble, le mutant de délétion PΔ1-144 du virus de Sendai ne fonctionne pas pour la transcription sur des nucléocapsides virales purifiées. Ceci suggère un rôle fonctionnel pour la région 1-144 (Curran *et coll.*, 1994). En fait, il s'agit d'un artefact : le mutant de délétion PΔ1-144 n'est plus capable d'empêcher l'auto-assemblage de N, et la forme assemblée de N séquestre PΔ1-144 qui n'est plus disponible pour la transcription. La polymérase est dégradée en l'absence de P, un fait symbolisé ici par des trous (Curran *et coll.*, 1994; Horikami *et coll.*, 1994).

B. En l'absence de N soluble, PΔ1-144 fonctionne pour la transcription (Curran *et coll.*, 1995b).

3) Comme je le propose dans la revue, il est probable que chez les ***Rubulavirus,*** **PNT** contienne des régions désordonnées. Après ce premier article, nous avons tout à fait les moyens conceptuels et expérimentaux de vérifier cette hypothèse. En particulier, avec les progrès en matière d'expression et de purification à haut débit de protéines recombinantes, il est tout à fait possible et souhaitable de purifier PNT issu de plusieurs *Rubulavirus* ayant un rôle modèle (par exemple SV5) ou un impact en santé humaine (par exemple, le virus des oreillons ou le virus paragrippal de type 2). Dans cette droite lignée, je pense qu'il serait judicieux d'exprimer aussi les modules PCT des mêmes virus, en vue de résoudre la structure tridimensionnelle de l'un d'entre eux au moins. Une telle approche parallèle sur plusieurs virus homologues maximiserait les chances de réussite, sans augmenter sensiblement l'effort de préparation des protéines.

ARTICLE N°2 : "SUBSTITUTION OF TWO RESIDUES IN THE MEASLES VIRUS NUCLEOPROTEIN RESULTS IN AN IMPAIRED SELF-ASSOCIATION"

RESUME

La substitution de deux résidus dans la nucléoprotéine du virus de la rougeole empêche son auto-assemblage en nucléocapside

La nucléoprotéine (N) du virus de la rougeole encapside l'ARN génomique viral en une nucléocapside hélicoïdale. Sa forte auto-association constitue un obstacle majeur à la détermination de sa structure par cristallographie aux rayons X. Nous avons exprimé chez *E. coli*, purifié et caractérisé un variant de N (NQD) qui a perdu sa capacité à former des pseudo-nucléocapsides, après substitution de deux résidus par des résidus polaires. Par des expériences d'immunoprécipitation, de dichroïsme circulaire, et de protéolyse limitée, nous montrons que cette nucléoprotéine conserve un repliement similaire au N original. De plus, elle se lie à la phosphoprotéine, et présente donc une certaine pertinence d'un point de vue biochimique. Nous présentons aussi des preuves que l'extrémité N-terminale de N se trouve à la surface de la nucléocapside. Au-delà de l'identification d'une région de N impliquée dans la polymérisation, nos résultats devraient faciliter les études structurales de N par cristallographie aux rayons X.

INTRODUCTION

La résolution de la structure tridimensionnelle des nucléoprotéines des *Mononegavirales* représente un défi, en particulier à cause de la taille et de l'inhomogénéité des polymères formés par N. Nous avons entrepris une approche de mutagenèse dirigée de N dans l'espoir d'obtenir une forme non polymérisée. Dans cette optique, un article de mutagenèse dirigée de N (Bankamp *et coll.*, 1996) a retenu notre attention. Celui-ci présente une délétion (celle de la région 189-239) qui abolit l'auto-assemblage de N tout en conservant sa liaison à P. Il s'agit d'un résultat très intéressant puisque N (ou du moins sa partie N-terminale, N_{CORE}) ne présente pas une organisation modulaire (un fait sur lequel les études existantes ne s'étendent guère). En effet, la quasi-totalité des délétions ou des substitutions d'acides aminés tentées jusqu'à présent produit un phénotype sans intérêt pour la cristallographie. Cependant, la délétion effectuée dans l'article originel (50 aa) risquait de provoquer des changements de conformation majeurs et délétères à la protéine. Nous avons donc effectué plusieurs substitutions *ponctuelles* dans cette région de N, au sein d'un motif conservé chez les *Paramyxovirinae*. L'une d'entre elles a conduit à un variant de N (NQD) au phénotype intéressant, qui fait l'objet du présent article.

Substitution of two residues in the measles virus nucleoprotein results in an impaired self-association

David Karlin[#], Sonia Longhi[#], and Bruno Canard*

Running title: Self-association of Measles virus nucleoprotein

Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Université Aix-Marseille I et II, ESIL, Campus de Luminy, 13288 Marseille Cedex 09, France

[#] contributed equally to this work

* to whom correspondence should be addressed

ESIL-CNRS-AFMB Case 925
163, avenue de Luminy
13288 Marseille Cedex 09
France
E-mail bruno@esil.univ-mrs.fr
Tel: (33) 4 91 82 86 44
Fax: (33) 4 91 82 86 46

Soumis à *Virology*

SUMMARY

The nucleoprotein (N) of measles virus encapsidates viral genomic RNA to form a helical nucleocapsid. Its strong self-association is a major hurdle in determining its high-resolution structure using X-ray crystallography. We report the bacterial expression, purification and characterization of a variant N that has lost its ability to form nucleocapsid-like structures after substitution of two residues by polar residues. Using immuno-precipitation, circular dichroism and limited proteolysis studies, we show that this nucleoprotein retains a folding similar to wild-type N. Furthermore, the variant N binds the phosphoprotein, indicating that it retains biochemical relevance. We also present evidence indicating that the N-terminus of N lies at the surface of the nucleocapsid. Beyond the identification of one region of N involved in self-association, our results should facilitate structural studies of N using X-ray crystallography.

KEY WORDS

Paramyxoviridae, *morbillivirus*, measles virus, nucleocapsid, nucleoproteins, point mutation, phosphoproteins, protein-protein interactions, binding sites, electron microscopy.

INTRODUCTION

Measles virus (MV), a member of the family *Paramyxoviridae,* is a major human pathogen (Griffin, 2001). Its non-segmented, negative-sense, single-stranded RNA genome is packaged by the viral nucleoprotein (N) within a helical nucleocapsid (Finch and Gibbs, 1970). Transcription and replication are carried out on this (N:RNA) complex by the polymerase (L) associated with the phosphoprotein (P) (reviewed in (Curran and Kolakofsky, 1999; Lamb and Kolakofsky, 2001; Sedlmeier and Neubert, 1998)). During genome replication, viral RNA is encapsidated by N concurrently to its synthesis (Gubbay *et al.*, 2001). N has the capacity to self-assemble on cellular RNA to form nucleocapsid-like particles in the absence of viral RNA and of any other viral protein (Spehner *et al.*, 1991; Warnes *et al.*, 1995). Therefore, a regulatory mechanism is necessary to prevent its illegitimate self-assembly in the absence of ongoing genomic RNA synthesis. Indeed, N is found in several conformations in infected cells. It is first synthesized as a soluble protein localized in the cytosol and also in the nucleus in the case of MV (Bohn *et al.*, 1990; Gombart *et al.*, 1993; Horikami and Moyer, 1995). Association of P with N prevents its illegitimate self-assembly, and also retains the soluble form of N in the cytoplasm (Curran *et al.*, 1995; Huber *et al.*, 1991; Spehner *et al.*, 1997). This N-P complex (called N°-P) is used by the polymerase to initiate encapsidation of neosynthesized genomic RNA (Curran *et al.*, 1995; Horikami *et al.*, 1992). The assembled form of N (called N^{NUC}) also forms complexes with P, either isolated (N^{NUC}-P) or bound to L (N^{NUC}-(P-L)), which are essential to RNA synthesis by the viral polymerase (Buchholz *et al.*, 1994; Curran, 1996; Curran, 1998; Ryan and Portner, 1990). The regions of N and P involved in N^{NUC}-P complex formation differ from those involved in N°-P (Curran *et al.*, 1995; De *et al.*, 2000; Harty and Palese, 1995; Horikami *et al.*, 1996; Karlin *et al.*, 2002; Liston *et al.*, 1997; Shaji and Shaila, 1999).

The structural and functional organization of MV N is detailed in Fig. 1. Deletion analysis and electron microscopy (EM) studies have shown that nucleoproteins of *Paramyxoviridae* are divided in two regions: an N-terminal moiety (herein called N_{CORE}), well conserved in sequence, (MV: aa 1-400), and a hypervariable, C-terminal moiety, N_{TAIL} (MV: aa 401-525) (Lamb and Kolakofsky, 2001). Within nucleocapsids, N_{TAIL} protrudes from the globular body of N_{CORE} (Heggeness *et al.*, 1981). N_{TAIL} contains the region responsible for N^{NUC}-P binding (aa 456-525 in MV N) and is consequently required for viral transcription (Bankamp *et al.*, 1996; Buchholz *et al.*, 1994; Liston *et al.*, 1997; Ryan *et al.*, 1993; Schoehn *et al.*, 2001).

Figure 1

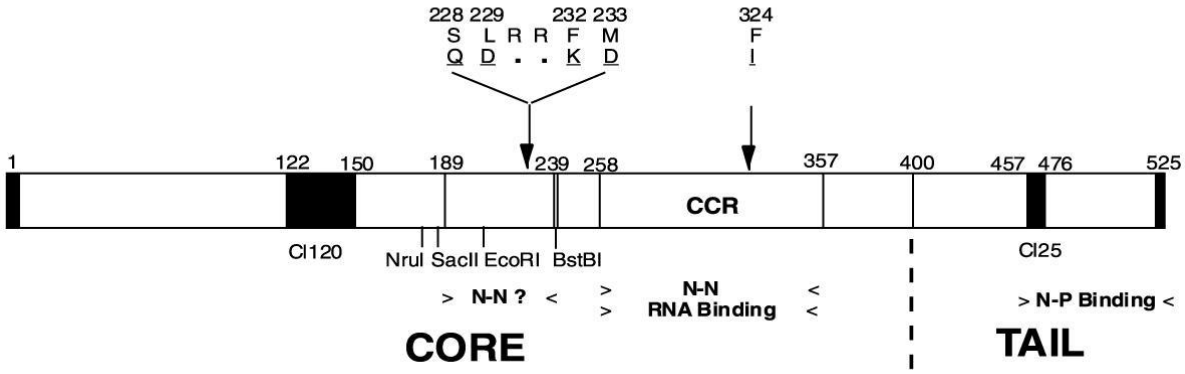

Fig. 1. Domain organization of MV nucleoprotein

N is divided into two regions: NCORE (aa 1-400) and NTAIL (aa 401-525), separated on the drawing by a dashed line below N. Numbering refers to the untagged N. The epitopes recognized by the monoclonal antibodies Cl120 and Cl25 are shaded. The N-terminal Flag and C-terminal hexahistidine tag in the recombinant NFlag-H6 are indicated by black bars at the corresponding termini of N. The N-N (boxed), N-P and RNA-binding regions are indicated below N. The positions targeted for mutagenesis are shown by arrows. Wild-type residues are shown in top position, whereas mutated residues are shown below. Additional unique restriction sites are also shown.

N_{CORE} contains all the regions necessary for self-assembly and RNA-binding, since nucleoproteins composed only of the core region can encapsidate neosynthesized RNA into nucleocapsid-like particles (Buchholz *et al.*, 1993; Curran *et al.*, 1993). Within N_{CORE}, deletion studies have identified regions involved in N-N interaction. The region spanning aa 258-357, called the Central Conserved Region (CCR), is well conserved in sequence, and mainly hydrophobic (Fig. 1). Several studies have shown that it constitutes one of the regions involved in self-association (Liston *et al.*, 1997; Myers *et al.*, 1997; Myers *et al.*, 1999), and in RNA binding (Kouznetzoff *et al.*, 1998; Myers *et al.*, 1997; Myers *et al.*, 1999). Beyond the CCR, another region of N-N interaction is found within the first 255 aa of *Paramyxoviridae* N (Myers *et al.*, 1997). Residues 189-239 of MV N could correspond to such a region, since deletion of this region leads to a nucleoprotein which can still bind P but has lost its ability to self-assemble (Bankamp *et al.*, 1996).

To date, no three-dimensional structure of a *Paramyxoviridae* nucleoprotein is known. A major hurdle to X-ray crystallography techniques is the strong self-assembly of N to form large nucleocapsids with a broad size distribution when expressed in heterologous systems such as mammalian cells (Spehner *et al.*, 1991), bacteria (Warnes *et al.*, 1995), or insect cells (Fooks *et al.*, 1993). Due to variable helical parameters, the recombinant or viral nucleocapsids are also difficult to analyze using EM coupled to image analysis. However, the low-resolution structure of the Rabies virus (RV, a *Rhabdoviridae*) nucleoprotein in its

assembled form has been determined using this method (Schoehn *et al.*, 2001). RV N is formed of two lobes, each lobe making contacts with the corresponding lobe of both the preceding and the following nucleoprotein. Functional and structural similarities between the nucleoproteins of *Rhabdoviridae* and *Paramyxoviridae* are well established. In particular, they share the same organization in two well-defined regions, N_{CORE} and N_{TAIL}, and in both families the CCR is involved in RNA-binding and self-assembly of N (Kouznetzoff *et al.*, 1998; Myers *et al.*, 1997; Myers *et al.*, 1999). However, they also have functional differences, such as the number of nucleotides bound per monomer (Lamb and Kolakofsky, 2001; Rose and Whitt, 2001). It is not known whether they share the same bilobal morphology.

In order to map further the regions of self-association of N and to obtain a nucleoprotein amenable to crystallization, we devised point substitutions of the CCR and of the region 189-239. One substitution impaired self-association of N, while retaining structural and functional characteristics of wild-type N.

RESULTS

When expressed in heterologous systems, N self-assembles into very large oligomers that are unsuitable for high-resolution structural studies, due to large sample heterogeneity. To circumvent this problem and to map N-N interactions, we have designed several mutants of the putative regions involved in self-assembly of N.

Design of mutant nucleoprotein genes

Oligomers of MV N expressed in bacteria do not dissociate in NaCl concentrations up to 1.5 M (unpublished data), suggesting that N-N interactions are not of ionic nature. One the regions responsible for these interactions is the CCR (aa 258-357) (Fig. 1), which contains numerous conserved hydrophobic residues (Liston *et al.*, 1997; Myers *et al.*, 1997; Myers *et al.*, 1999). Mutants of the SeV N CCR in which F324 is replaced by either L or I exhibit defects in self-assembly (Myers *et al.*, 1997). Since this phenylananine residue is conserved among all *Morbillivirus* and *Paramyxovirus* nucleoproteins (not shown), we replaced it by an isoleucine in MV N to yield $NI324_{Flag\text{-}H6}$.

Another putative region involved in the N-N interaction is formed by aa 189-239 (Bankamp *et al.*, 1996). The secondary structure prediction of this region in *Paramyxovirinae* N shows the presence of two α-helices (Fig. 2). When looking at the second helix (aa 213-238), two types of conserved hydrophobic residues can be found: hydrophobic amino acids spaced with a periodicity of 4 or 5 residues, which may correspond to residues located at the inner side of an amphipatic helix, and hydrophobic residues (shown in bold and underlined in Fig. 2), which do not respect this periodicity and which would be on the exposed side of such a helix. We reasoned that they might mediate the N-N interaction. Consequently, we designed mutants in which two or four of these hydrophobic residues are replaced with polar residues. In the double mutant $NQD_{Flag\text{-}H6}$, residues SL (228-229) were changed to QD; in the double mutant $NKD_{Flag\text{-}H6}$, FM (232-233) were replaced by KD; and in the tetra mutant $NQDKD_{Flag\text{-}H6}$, all four residues were replaced by QD and KD (see Fig 1 and 2).

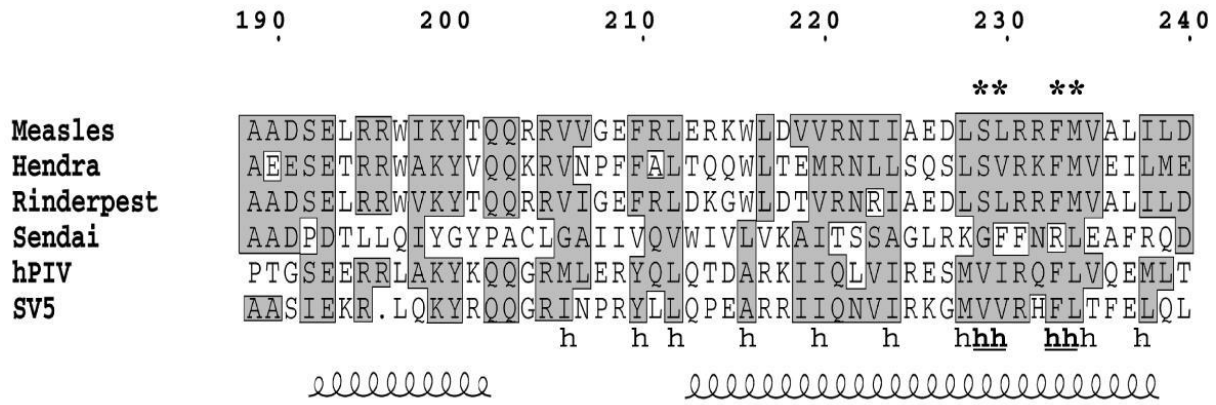

Fig. 2. Region of self-association of N targeted for mutagenesis (aa 189-239)

Multiple sequence alignment of the N proteins of Paramyxoviridae. Accession numbers are as follow: hPIV (human parainfluenza virus type 4A): P17240. Measles: P04851. Hendra: 089339. Rinderpest: P37708. Sendai: P04858. SV5 (Simian virus type 5): Q88435.

Predicted secondary structure elements (α-helices) are shown below the alignment. Homologous residues (plurality of 65%) are shown in boxes. Highly hydrophobic positions (i.e. positions corresponding to at least 6 hydrophobic, or 5 aromatic, or 4 hydrophobic + 2 homologous residues) are marked by a "h". Positions targeted for mutagenesis are indicated by a "h" shown in bold and underlined. Asterisks above the multiple sequence alignment indicate positions of amino acid substitutions.

Expression and purification of native and substituted nucleoproteins

The N gene of MV contains numerous arginine codons AGA that are used with a very low frequency in *E. coli*. Therefore, we co-expressed the plasmid pUBS520, which supplies the rare corresponding tRNA (Brinkmann *et al.*, 1989; Karlin *et al.*, 2002) with pET21a/$N_{Flag\text{-}H6}$ (and all mutant-encoding plasmids derived from it). Most $N_{Flag\text{-}H6}$ produced in bacteria grown at 37°C was found in the soluble fraction (Fig. 3A, lane SN*)*. $N_{Flag\text{-}H6}$ was purified to homogeneity (> 95 %) in two steps: Immobilized Metal Affinity Chromatography (IMAC) and gel-filtration (Fig. 3A, lanes IMAC and GF). The bands migrating slightly faster than N (Fig. 3A, lane GF) are degradation products of N as confirmed using Western Blot (WB) with anti-N antibodies (not shown).

On the other hand, none of the mutants produced in bacteria grown at 37°C were found in the soluble fractions. When expressed at 17°C, $NKD_{Flag\text{-}H6}$ and $NQDKD_{Flag\text{-}H6}$ were not soluble and were degraded in C-terminal, as seen using WB with anti-Flag antibodies (not shown). $NI324_{Flag\text{-}H6}$ was poorly soluble (Fig. 3B, lane SN) and was purified to about 30 % homogeneity in one IMAC step (Fig. 3B, lane IMAC). $NQD_{Flag\text{-}H6}$ was soluble (about 15 %) (Fig. 3C, lane SN) only in the presence of non-ionic detergents, and was purified to homogeneity (> 95 %) in one step by IMAC (Fig. 3C, lane IMAC).

Figure 3

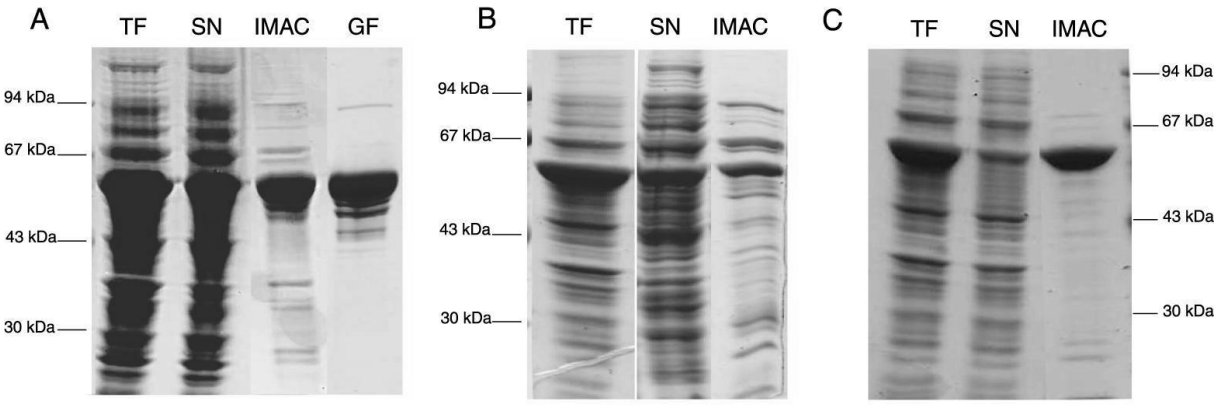

Fig. 3. Purification of the nucleoproteins

Coomassie blue staining of a 10% SDS-PAGE. Purification of N **(A)**, NI324 (**B)** and NQD (**C)**. TF: bacterial lysate (total fraction). SN: clarified supernatant (soluble fraction). IMAC: eluent from Immobilized Metal Affinity Chromatography. GF: eluate from gel filtration.

RNA content and molecular weight and of native and mutant nucleoproteins

We examined the ratio between the Optical Densities (OD) of purified nucleoproteins at 280 nm and at 260 nm, which give indications on their RNA content. The theoretical OD_{280}/OD_{260} ratio for N devoid of RNA is 1.4, whereas it is 0.9 for the (N:RNA) complex containing 6 ribonucleotides per N monomer (see under Material and Methods). The ratio measured for the purified recombinant $N_{Flag\text{-}H6}$ is 0.9, suggesting the presence of RNA within the nucleoprotein. On the other hand, the ratio measured for $NQD_{Flag\text{-}H6}$ (1.4) indicates that it contains little or no RNA.

We performed a gel-filtration analysis on $N_{Flag\text{-}H6}$, $NI324_{Flag\text{-}H6}$ and $NQD_{Flag\text{-}H6}$. They all eluted in the dead volume of the gel filtration column (not shown). This indicates a size greater than 1300 kDa, which corresponds either to unspecific aggregates or to high-degree multimers of N (58 kDa). EM was used to analyze the $NQD_{Flag\text{-}H6}$ aggregates further.

NQD no longer forms herringbone structures

N produced in MV-infected-cells forms nucleocapsids with a typical herringbone structure that can be visualized by EM (Finch and Gibbs, 1970; Nakai *et al.*, 1969). EM analysis of purified $N_{Flag\text{-}H6}$ reveals the presence of typical nucleocapsid-like, herringbone structures ranging in size from 18 to 48 nm (Fig. 4A, arrows). However, most of the nucleocapsid-like particles appear as rings having a diameter of 18 nm (Fig. 4A, feathered arrows), corresponding to shorter nucleocapsids seen perpendicularly to their axis (such a ring can be seen in (Warnes *et al.*, 1995), Fig. 4B).

On the other hand, no herringbone structure is visible on EM images of $NQD_{Flag\text{-}H6}$ (Fig. 4B), which forms regular blobs of smaller average size (10 nm). Thus, the assembly of NQD in

native-like polymers is disrupted. This might arise from *i*) disruption of the structure of the individual subunit of the polymer, i.e. major structural changes in NQD *ii*) non specific aggregates of the intact monomer and *iii*) a different, though specific self-association mode of the intact subunits. To distinguish between these possibilities, we explored the structure of NQD further using limited proteolysis and circular dichroism.

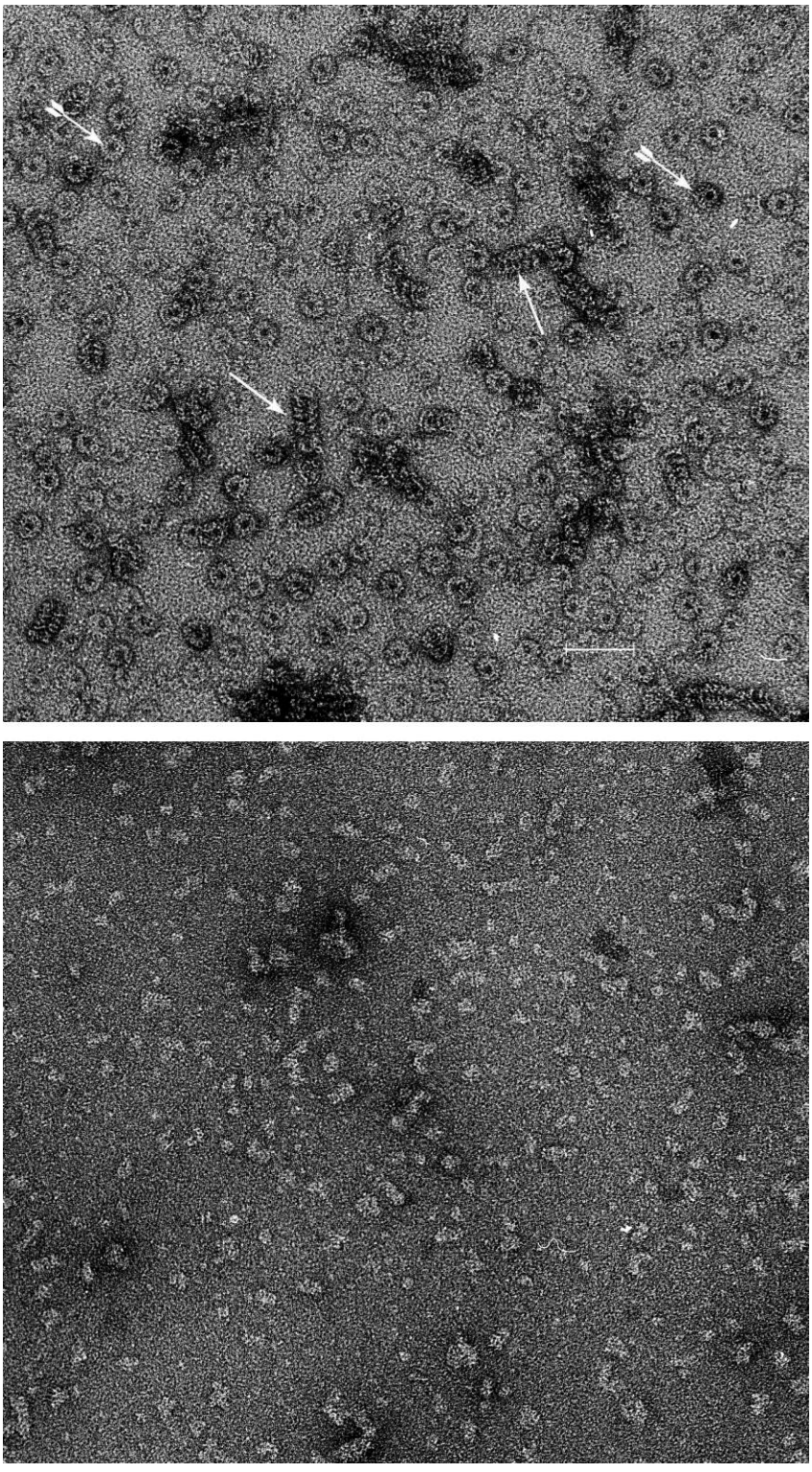

Fig. 4 Comparison of the association states of N and NQD by negative-staining EM

Negative staining electron micrographs. (A) $N_{Flag\text{-}H6}$. Nucleocapsid-like, herringbone structures are indicated by arrows. Rings (corresponding to short nucleocapsids seen perpendicularly to their axis) are indicated by feathered arrows.

(B). $NQD_{Flag\text{-}H6}$ forms smaller, regular aggregates. The bar represents 50 nm.

A new site sensitive to proteolysis is exposed in NQD

Limited proteolysis is a powerful tool to investigate structural changes caused by a mutation in a protein (Hubbard, 1998; Hubbard, 2001). To establish the proteolytic pattern of N, we incubated $N_{Flag\text{-}H6}$ in the presence of increasing concentrations of trypsin, which cleaves specifically after Arg or Lys residues. At low trypsin concentrations, $N_{Flag\text{-}H6}$ is digested to a fragment of about 45 kDa (Fig. 5A, lanes 0.25 and 0.5). Upon increasing trypsin concentrations, a highly resistant fragment of 43 kDa is obtained (Fig. 5A, lane 2.5 and 5). The 43 kDa fragment eluted in the void volume of the gel filtration column (not shown), indicating that it is either aggregated or still polymerized. A WB with an anti-Flag antibody (Fig. 5B) showed that this fragment contains the N-terminal Flag epitope, and is thus composed of the N-terminal moiety of N, N_{CORE}. On the other hand, no fragment corresponding to the remaining 15-17 kDa (N_{TAIL}) was detected, indicating that it is hypersensitive to proteolysis and entirely degraded. Indeed, purified N is partially degraded in its C-terminal (see Fig. 5B), as already reported for MV N purified from virions (Giraudon *et al.*, 1988). The 43 kDa polypeptide purified by gel-filtration was submitted to mass spectrometry, which indicated that it was composed of fragments with molecular weights (MW) ranging from 40 to 45 kDa. From their precise MWs, we deduced that in some fragments trypsin had cleaved off the first or the second Lys residue within the Flag, consistent with the decrease in WB signal (Fig 5B, lane 2.5). We estimated that these fragments corresponded to N-terminal fragments with C-termini ranging from aa 362 to 404 in the sequence of N. In a second, independent experiment, the major trypsin cleavage site corresponding to the highly resistant fragment was located more precisely to Arg 392 (not shown). Thus, after digestion of the hypersensitive tail until around aa 404, trypsin can further digest N until aa 392, and even as far as aa 362 in the presence of the highest enzyme/protein ratio (1:50) used here.

Figure 5

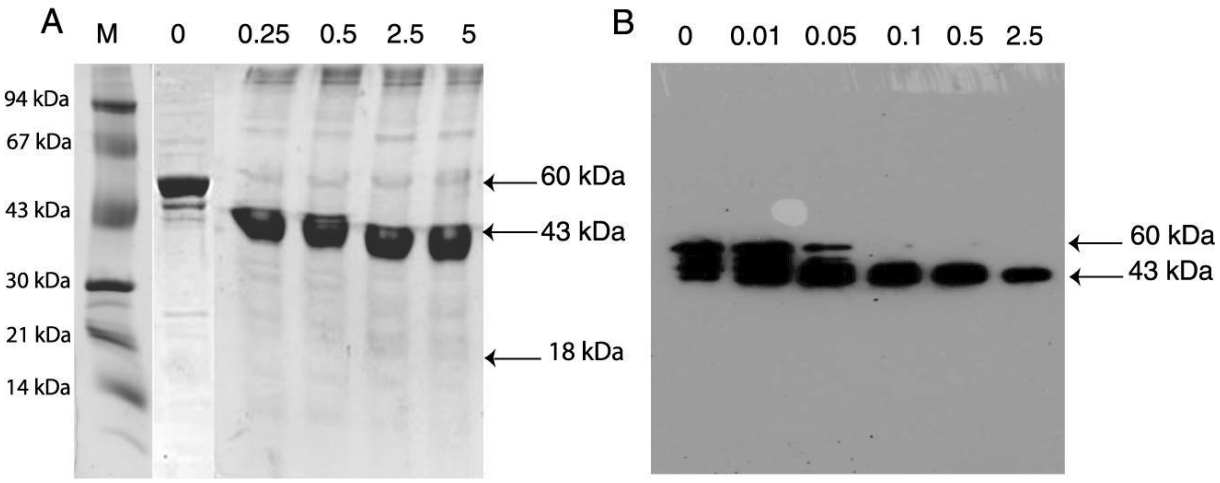

Fig. 5. Limited proteolysis of N

The extent of digestion of N (2 mg/ml) at various trypsin concentrations was monitored on a 15% SDS-PAGE. M = Markers. The arrows shows undigested $N_{Flag\text{-}H6}$ (60 kDa) and the 43 kDa fragment resistant to proteolysis. A Coomassie Blue staining. B. Western Blot with an anti-Flag (aFlag) antibody.

Figure 6

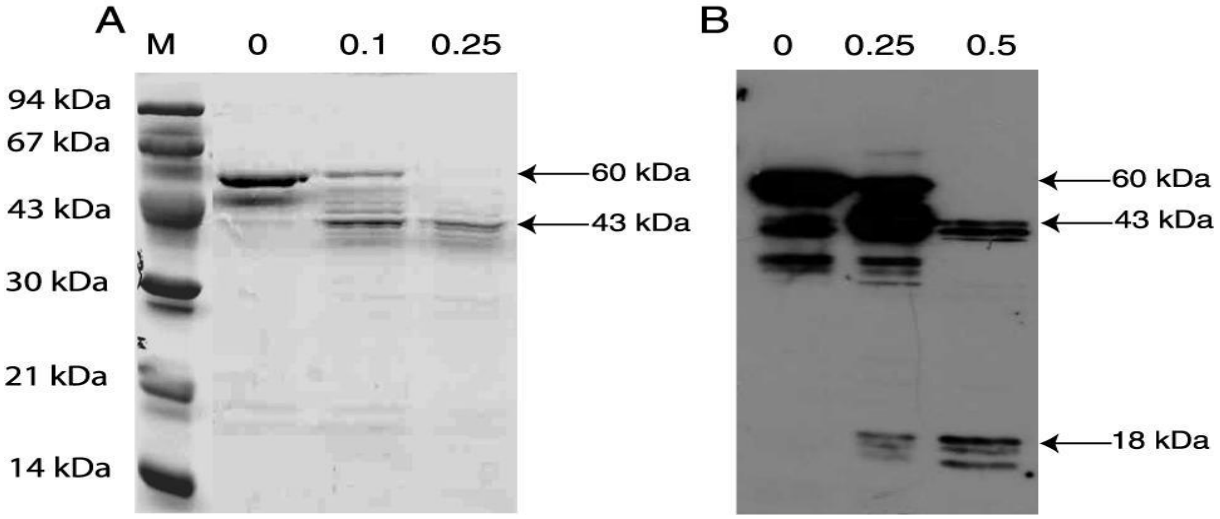

Fig. 6. Limited proteolysis of NQD

The extent of digestion of $NQD_{Flag\text{-}H6}$ (0.2 mg/ml) at various trypsin concentrations (µg/ml) was monitored. M = Markers (94, 67, 43, 30, 20, 14 kDa). The arrows show undigested $NQD_{Flag\text{-}H6}$ (60 kDa), the 43 kDa fragment resistant to proteolysis and a 18 kDa additional fragment. Digestion was monitored on a 15% SDS-PAGE. A Coomassie Blue Staining B. Western Blot with an anti-Flag (aFlag) antibody.

Like N, purified $NI324_{Flag\text{-}H6}$ and $NQD_{Flag\text{-}H6}$ were partially cleaved in their C-terminal domain (Fig. 6A and 6B, lane 0). The proteolytic pattern of $NI324_{Flag\text{-}H6}$, monitored using WB with an anti-Flag antibody, was similar to that of N, yielding a 43 kDa resistant core (not shown). Proteolysis of $NQD_{Flag\text{-}H6}$ at low trypsin concentrations also yielded a 43 kDa core (Fig. 6A, lane 0.25). However, upon increasing trypsin concentrations, and unlike wild-type N, this 43 kDa fragment was cleaved to yield an additional fragment of about 18 kDa (Fig. 6B, lane 0.5). The 18 kDa fragment, still containing the N-terminal Flag epitope, results from a cleavage at around aa 150. Thus, a new trypsin-sensitive site that is not accessible in wild-type nucleocapsids is exposed in the aggregates formed by NQD.

The circular dichroism (CD) spectra of N and NQD are similar

NQD has the same organization than N, being divided into a C-terminal tail hypersensitive to proteolysis and a N-terminal core. Whether the secondary structures of both nucleoproteins are similar was investigated using CD.

Fig. 7 shows the CD spectra of $N_{Flag\text{-}H6}$ (full circles) and $NQD_{Flag\text{-}H6}$ (open circles) recorded in 10 mM NaP pH 7.0. The spectrum of $N_{Flag\text{-}H6}$ shows a high positive ellipticity around 195nm and two minima at 208 and 222 nm. This spectrum is typical of proteins with a predominant α-helical content. Deconvolution of the spectrum using the algorithm CDNN (Dalmas *et al.*, 1994) gives a helical content of about 35%, in good agreement with the PHD secondary structure prediction (Rost, 1996) of 33.5 % helix (not shown). The deconvolution gives a low β-structure content (about 12%).

The spectrum of $NQD_{Flag\text{-}H6}$ is almost superimposable to that of $N_{Flag\text{-}H6}$, with the difference that the two minima at 208 and 222nm are less pronounced. Deconvolution of the spectrum of $NQD_{Flag\text{-}H6}$ also gives a helical content of about 35% and a low β-structure content.

These results suggest that the amino acid substitution in $NQD_{Flag\text{-}H6}$ has not affected its secondary structure significantly.

Figure 7

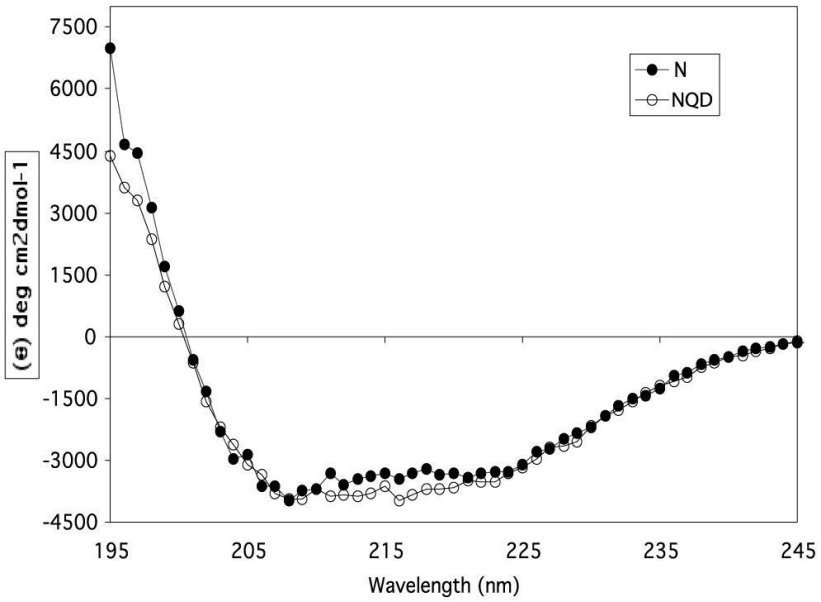

Fig. 7. Circular dichroism (CD) spectra of N and NQD

Full circles: spectrum of $N_{Flag\text{-}H6}$. Open circles: spectrum of $NQD_{Flag\text{-}H6}$. The spectra were recorded in 10 mM sodium phosphate buffer, pH 7.0 (see under Materials and Methods).

The same epitopes are exposed on N and NQD

To determine whether epitopes exposed on the surface of the native nucleoprotein were still exposed on NQD, we immunoprecipitated N and NQD from bacterial lysates using the anti-N monoclonal antibodies (mAbs) Cl120 and Cl25 (Buckland *et al.*, 1989; Giraudon *et al.*, 1988), directed against the regions 122-150 and 457-476 of N, respectively (see Fig. 1). Both antibodies recognize linear epitopes (i.e. they are not conformational antibodies). The substitution in NQD (aa 228-229) lies outside the epitopes recognized by Cl25 and Cl120, and accordingly, both antibodies recognize NQD on western blots (not shown). We also used the mAbs aH6 and aFlag, directed against the hexahistidine epitope and the Flag epitope, respectively. As expected, $N_{Flag\text{-}H6}$ is precipitated by Cl120 and by Cl25 (Fig. 8, left panel, lanes Cl25 and Cl120), but bacterial proteins from a control lysate are not (Fig. 8, lane pET), confirming the specificity of the precipitation. $N_{Flag\text{-}H6}$ is also precipitated by aFlag and aH6, indicating that both fusion tags are accessible on the surface of $N_{Flag\text{-}H6}$ (Fig. 8, lanes Flag and H6). The possibility that N could interact non-specifically with mouse antibodies is ruled out by the fact that untagged N is not precipitated by aH6 (Fig. 8, lane N).

As shown in Fig. 8, right panel, $NQD_{Flag\text{-}H6}$ is also precipitated by antibodies Cl120, Cl25, aH6 and aFlag, indicating that all the corresponding epitopes are exposed on the surface of $NQD_{Flag\text{-}H6}$.

Figure 8

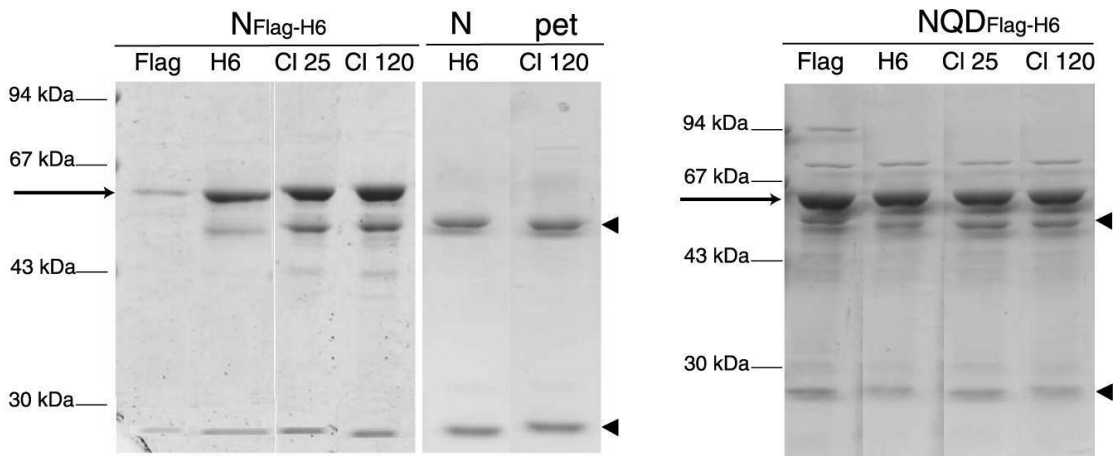

Fig. 8. Immunoprecipitation studies on N and NQD.

The immunoprecipitation of N and NQD using various antibodies was monitored on a 10% SDS-PAGE stained by Coomassie blue.

Left panel.

Immunoprecipitation of $N_{Flag\text{-}H6}$ by the mAbs aFlag, aH6, Cl25 and Cl120. Lane N: immunoprecipitation of the untagged N by aH6. Lane pET: immunoprecipitation by Cl120 of a control bacterial lysate transformed with the parental (pET21a) plasmid. The same result was obtained with Cl25 (not shown).

Right panel.

Immunoprecipitation of $NQD_{Flag\text{-}H6}$ by the mAbs aFlag, aH6, Cl25 and Cl120 respectively.

Arrowheads indicate the migration of light and heavy chains of antibodies, which are visible on the gel around 25 and 55 kDa respectively. Arrows correspond to native and mutant nucleoproteins.

NQD binds P

As discussed above, N^{NUC} binds P (Huber *et al.*, 1991). To determine whether NQD also binds P, we tested the ability of bacterial lysates expressing tagged N or NQD to pull down untagged P on an IMAC resin. Since N and P are expressed separately, N is produced in its self-assembled form N^{NUC} and the interaction tested is the N^{NUC}-P interaction. The IMAC resin selectively pulled down N_{H6} expressed alone but not the untagged P expressed alone (Fig. 9, lanes N and P). As expected, when both lysates were mixed, the IMAC resin selectively pulled down N_{H6} together with a protein of the MW expected for P (Fig. 9, lane N+P). The identity of the precipitated protein was confirmed using a P protein tagged with an N-terminal Flag (P_{Flag}) (not shown). To rule out the possibility that the hexahistidine tag of N_{H6} could be responsible for the precipitation of P, we used the hexahistidine-tagged N-terminal domain of the MV phosphoprotein (PNT_{H6}, (Karlin *et al.*, 2002)) as a control, since PNT does not bind N^{NUC} (Liston *et al.*, 1995). As expected, PNT_{H6} did not pull down N^{NUC} (not shown).

As shown in Fig. 9, lane NQD+P, NQD also selectively pulled down P, indicating that it binds P.

Figure 9

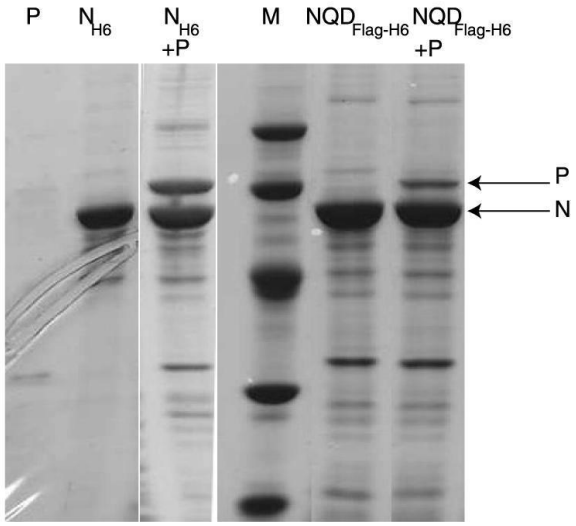

Fig. 9. Binding of N and NQD to P

Precipitation by an IMAC resin of untagged P (P), NH6 (N), NH6 + untagged P (N+P), $NQD_{Flag\text{-}H6}$ (NQD), and $NQD_{Flag\text{-}H6}$ + untagged P (NQD+P). Coomassie blue staining of a 10% SDS-PAGE. M: markers (94, 67, 43, 30, 20, 14 kDa). The arrows indicate the migration of N and P.

Taken together, our results suggest that NQD is similar to the native N immunogenically (as it is recognized by anti-N monoclonal antibodies), structurally (since its secondary structure is comparable to that of N and it still possesses a 43 kDa core), and functionally (since it binds P), but that its self-association is impaired, as seen using EM.

DISCUSSION

We have purified and characterized the bacterially expressed nucleoprotein of MV, with an N-terminal, 8 aa hydrophilic Flag (Brizzard *et al.*, 1994) and a C-terminal hexahistidine tag ($N_{Flag\text{-}H6}$). N expressed in bacteria is morphologically and immunogenically indistinguishable from virion-extracted N (Warnes *et al.*, 1995). Likewise, the tagged $N_{Flag\text{-}H6}$ forms typical herringbone nucleocapsid-like structures, and reacts with two monoclonal antibodies directed against different regions of N, indicating that the N-terminal Flag does not prevent its proper folding and assembly. This is noteworthy, since N_{CORE} is very sensitive to deletions (Bankamp *et al.*, 1996; Buchholz *et al.*, 1993; Curran *et al.*, 1993; Myers *et al.*, 1997; Myers *et al.*, 1999). In particular, deletions in the N-terminus of N (aa 1-82) led to the formation of unstructured aggregates (Buchholz *et al.*, 1993). However, in agreement with our results, the introduction of a 40 aa tag fused to the N-terminus of Newcastle Disease Virus (NDV) N does not impede folding and self-assembly (Errington and Emmerson, 1997).

Limited proteolysis of N yields a 43 kDa N-terminal fragment corresponding to N_{CORE}, while the C-terminal N_{TAIL} is entirely degraded, confirming previous reports (Giraudon *et al.*, 1988; Heggeness *et al.*, 1981). The lack of monoclonal antibodies generated against the very N-terminus of N, and the fact that it is not attacked by proteases suggested that it was hidden inside the nucleocapsid structure (Gill *et al.*, 1988; Liston *et al.*, 1997). However, the N-terminal Flag of $N_{Flag\text{-}H6}$ is accessible to an anti-Flag antibody. This suggests that the N-terminus of N lies on the surface of the nucleocapsid, in agreement with the finding that relatively large N-terminal tags cause no steric hindrance to the folding and assembly of N (see above).

The mapping and the structural determinants of N self-polymerization are poorly understood. A MV nucleoprotein in which aa 189-239 are deleted does not self-assemble but still binds P (Bankamp *et al.*, 1996), suggesting that this region is important for nucleocapsid formation. In this study we have examined the effect of substitutions in this region. The mutant NQD, in which two conserved residues (S,L) have been replaced by polar residues (Q,D) at position 228-229 was of particular interest. We have also designed a mutant of the CCR (called NI324) in which F324 (conserved among *Paramyxovirus* and *Morbillivirus* N) is replaced by I, based on the observation that the corresponding SeV NI324 does not self-assemble in mammalian cells [Myers, 1997 #75].

Remarkably, MV NI324 is stable in bacteria, contrary to SeV NI324, which is unstable in mammalian cells (Myers *et al.*, 1997). NI324 and NQD are both less soluble in *E. coli* than N, suggesting that they expose new hydrophobic surfaces to the solvent. Using gel-filtration, we showed that NI324 and NQD form high-degree multimers.

The pattern of proteolysis of NI324 is indistinguishable from that of N. Furthermore, it is still multimeric, as opposed to the SeV NI324 mutant (Myers et al., 1997), suggesting that

the phenylalanine 324 is not the sole self-assembly determinant of MV N. On the other hand, the multimers formed by NQD are not nucleocapsid-like structures, rather smaller, regular aggregates, as seen using EM. To determine if this was due to a major structural change in the subunit of the polymer, we performed direct structural studies on NQD. NQD still has the typical 43 kDa core of wild-type N. However, a new trypsin-sensitive site that is not accessible in wild-type nucleocapsids is exposed in the aggregates formed by NQD. The circular dichroism spectrum of NQD is very similar to that of N, suggesting that the substitution has not markedly affected its secondary structure. We could not distinguish NQD from N immunogenically either, using antibodies directed against two epitopes located in the core and in the tail of N, and both exposed on the outside of the nucleocapsid (Giraudon et al., 1988). Furthermore, in agreement with (Bankamp et al., 1996), NQD retains part of its biochemical relevance, as it still binds P. Together, these results indicate that NQD has retained a folding similar to that of N. It is not possible to ascertain at this stage whether the SL(228-229)>QD mutation disrupted self-association of N partially, or if NQD is aggregating randomly.

Recently, (Schoehn *et al.*, 2001) have shown that there are 4 strong N-N binding sites on each N molecule of RV (a *Rhabdoviridae*), 2 for each neighbor. Consequently, they pointed out that disruption of 2 N-N binding sites would probably be necessary for crystallization of RV N. The number of N-N binding regions in *Paramyxoviridae* N proteins is not known, but they contain at least 2 such regions. One is located within the CCR and another within aa 1-255 (Liston *et al.*, 1997; Myers *et al.*, 1997; Myers *et al.*, 1999), likely in the region 189-239 (Bankamp *et al.*, 1996); this study.

Other deletions of N that lead to the formation of aggregates have been reported (Bankamp *et al.*, 1996; Buchholz *et al.*, 1993; Myers *et al.*, 1997; Myers *et al.*, 1999). Deletion of the whole region 189-239 prevents multimerization of N (Bankamp *et al.*, 1996), whereas a point substitution in this region yields another form of N-N interaction. The large deletion (50 aa) might cause structural damages in N. However, none of the resulting nucleoproteins has been characterized from a structural point of view.

Structural studies of N proteins at the atomic level are mainly limited by the self-polymerization problem, as well as by the fact that N package unspecific RNA when expressed in the absence of ongoing viral replication. All mutants with an impaired N-N association characterized so far do not package RNA, suggesting that the RNA-binding site of N is at least in part formed by the self-assembly of N (Kouznetzoff et al., 1998; Myers et al., 1997; Myers et al., 1999). Likewise, NQD contains little or no RNA, as indicated by its OD_{280}/OD_{260} ratio.

In conclusion, we have mapped a region critical for self-polymerization of MV N. Substitution of two amino acids in this region to yield NQD led to another form of N-N interaction. The structure of the monomer within the aggregates formed by NQD is not altered. NQD does not

contain RNA, and retains P-binding activity. The aggregates formed by NQD are homogenous enough to make NQD an attractive candidate for crystallization studies. The strategy employed here might be applied to other viral proteins for which macromolecular assemblies impede structural studies.

Material and Methods

Bacterial strains, chemicals and antibodies

Bacterial strains and chemicals were as described in (Karlin *et al.*, 2002). We used two anti-N monoclonal antibodies (mAbs), Cl 25 and Cl 120 (Buckland *et al.*, 1989; Giraudon *et al.*, 1988), kindly provided by D. Gerlier.

Cloning of the P coding region

The plasmids pET21a/P_{H6} and pET21a/P encode the phosphoprotein of measles virus (strain Edmonston B), with or without an N-terminal hexahistidine tag respectively (Karlin *et al.*, 2002). We constructed a third plasmid, pET21a/P_{Flag}, encoding the P protein with a Flag sequence (DYKDDDDK) (Brizzard *et al.*, 1994) fused to its N-terminus. pET21a/P was digested with *Nde*I and *BamH*I, dephosphorylated, and ligated to a double-stranded, phosphorylated synthetic oligonucleotide (5' TATGGATTATAAAGACGATGACGACAAAGCAGAAGAGCAGGCACGCCATGTCAAAAACG GACTCGAGTGCA 3' annealed to 5' CTCGAGTCCGTTTTTGACATGGCGTGCCTGCTCTTCTGCTTTGTCGTCATCGTCTTTATA ATCCA 3') to yield pET21a/P_{Flag}.

Cloning of the N coding region

The plasmid pSC6/N encodes the N gene of the measles virus (strain Edmonston B) (Radecke *et al.*, 1995). N was cloned into the expression vector pET21a (Novagen) by PCR using pSC6/N as a template. Forward primer (5' GGCCCATATGGCCACACTTTTAAG GAG C 3') was designed to introduce an *Nde*I site fused at the N-terminus of N, while the reverse primer (5' CCGGGAATTCTTAGTCTAGAAGATTTCTGTC 3') introduced an *EcoR*I site after the N stop codon. The PCR product was digested with *Nde*I and *EcoR*I and purified. The resulting fragment was ligated between the *Nde*I and *EcoR*I sites of pET21a to yield pET21a/N.

2 plasmids for the expression of tagged, recombinant N proteins were generated from pET21a/N: pET21a/NFlag which encodes N with a Flag sequence fused to its N-terminus, and pET21a/NFlag-H6 in which an hexahistidine tag is fused to the C-terminus of NFlag. pET21a/N was digested with *Nde*I and *BamH*I, dephosphorylated, and ligated to a double-stranded, phosphorylated synthetic oligonucleotide (5' tATGGATTATAAAGACGATGACGACAAAgccacacttttaaggagcttagcattgttcaaaagaaacaaggaca aaccacCcattacCtcGG 3' annealed to 5'GATCcCgaGgtaatgggtggtttgtccttgtttcttttgaacaatgctaagctccttaaaagtgtggcTTTGTCGTCATC GTCTTTATAATCCA 3') to yield pET21a/N_{Flag}.

A hexahistidine tag was fused to the C-terminus of N_{Flag} by PCR using pET21a/N_{Flag} as a template, forward primer (5' attgcaatgcatactactgaggac 3'), and reverse primer

(5'ggccGAATTCttaGTGATGGTGATGGTGATGgtctagaagatttctgtcattgtacactat 3'). The PCR product was digested with *Nsi*I and *EcoR*I and purified. The resulting fragment was ligated between the *Nsi*I and *EcoR*I sites of pET21a/N_{Flag} to yield pET21a/$N_{Flag-H6}$. The sequences of all coding regions were confirmed by sequencing.

Construction of the plasmids encoding mutant nucleoproteins

The region between aa 189-239 of N, that we targeted for mutations, corresponds to bp 565-717 in the sequence of the N open reading frame (ORF). To allow the rapid introduction of mutations in this region, we constructed a cassette where several additional unique restriction site were inserted without changing the wild-type amino-acid sequence (silent sites). pET21a/$N_{Flag-H6}$ was used to derive a construct in which the $N_{Flag-H6}$ ORF possesses 4 additional unique restriction sites (*Nru*I, *Sac*II; *EcoR*I, *BstB*I) (see Fig. 1). First, the 3' proximal *EcoR*I site of pET21a/$N_{Flag-H6}$ was eliminated by digesting pET21a/$N_{Flag-H6}$ sequentially with *EcoR*I and Mung Bean nuclease, followed by religation. Then the silent mutations were introduced by several PCR steps, using either synthetic oligonucleotides alone, or PCR-amplified DNA fragments combined with synthetic oligonucleotides as primers. The resulting construct contains the following unique sites within the region targetted for mutations: *Bsu36*I (502); *Nru*I (557), *Sac*II (584); *EcoR*I (646), *BstB*I (680), *Xma*I (754) (numbers in brackets correspond to their nucleotide position in the $N_{Flag-H6}$ ORF).

Mutants NQD, NKD and NQDKD were then constructed by inserting phosphorylated, double-stranded synthetic oligonucleotides containing mutations between the *BstB*I and *Xma*I sites. Codons TCC TTA encoding SL (228-229) were replaced with CAG GAT, coding for QD, to yield pET21a/$NQD_{Flag-H6}$. In pET21a/$NKD_{Flag-H6}$, codons TTC ATG encoding FM (232-233) were replaced with AAA GAC, coding for KD. pET21a/$NQDKD_{Flag-H6}$ contains both mutations.

The phenylalanine residue encoded at position 970 of the N ORF was replaced by an isoleucine using recursive PCR. The first round of PCR, using pET21a/$N_{Flag-H6}$ as template, yielded two PCR products, using 1) forward primer 5' AGTGCAAGACCCTGAGGGATTCAAC 3' with reverse primer 5' AGGGTAGGATCCTGCACTAATCTTGTTCTGAA 3' and 2) forward primer 5' TTCAGAACAAGATTAGTGCAGGATCCTACCCT 3' with reverse primer 5' CCTCAGTAGTATGCATTGCAATCTC 3'. The two PCR products were used as reciprocal primers for a subsequent round of PCR, and the final product was ligated between the *Bsu36*I and *Nsi*I sites of pET21a/$N_{Flag-H6}$ to yield pET21a/$NI324_{Flag-H6}$. The sequences of all coding regions were confirmed by sequencing.

Expression of P and N

E. coli BL21(DE3) transformed with either pET21a/$N_{Flag-H6}$ or pET21a/$P_{Flag-H6}$ in the presence of pUBS520 (Brinkmann *et al.*, 1989) were grown and induced as described for pET21a/P_{H6} (Karlin *et al.*, 2002).

Expression of the variant nucleoproteins

E coli BL21(DE3) transformed with a mutant-encoding plasmid (pET21a/$NQD_{Flag\text{-}H6}$, or pET21a/$NQDKD_{Flag\text{-}H6}$, or pET21a/$NKD_{Flag\text{-}H6}$, or pET21a/$NI324_{Flag\text{-}H6}$) and with pUBS520 (Brinkmann *et al.*, 1989) was grown in Luria-Bertani medium containing 100 µg/mL ampicilin and 50 µg/mL kanamycin. At an OD_{600} of 0.5, the cells were incubated on ice for 2 hrs, then 2% ethanol and 50 µM IPTG were added, and the cells were grown at 17°C for 23 hrs. The induced cells were harvested, washed and collected by centrifugation. The resulting pellets were frozen at –20°C.

Purification of N

Starting from a 1 l cell culture, the pellet containing $N_{Flag\text{-}H6}$ was resuspended in buffer A (50 mM sodium phosphate pH 8.0, 300 mM NaCl) supplemented with 20 mM imidazole, 100 µg/ml lysozyme, 1 µg/ml DNAse I, 125 µl protease inhibitor cocktail (Sigma). After a 20 min incubation on ice, the cells were disrupted by sonication (using a 750 W sonicator for 4 cycles of 50 s at 50 % power output). The lysate was clarified by centrifugation at 30 000g for 30 min at 4°C.

The clarified supernatant was incubated for 1 hr at 4°C with gentle shaking with 4 ml Chelating Sepharose Fast Flow Resin preloaded with Ni^{2+} ions (Amersham Pharmacia Biotech), previously equilibrated in buffer A containing 10 mM imidazole. The resin was washed with buffer A containing 20 mM imidazole, and N was eluted in buffer A containing 250 mM imidazole. Eluates were analyzed by SDS-PAGE for the presence of N. The fractions containing N were combined, loaded onto a Superdex 200 HR 10/30 column (Amersham Pharmacia Biotech) and eluted with buffer A. The protein was concentrated with Centripreps 30 (Millipore).

Purification of NI324

The purification of $NI324_{Flag\text{-}H6,}$ using IMAC, was identical to that of $N_{Flag\text{-}H6}$. Analytical gel filtration was carried out as described above.

Purification of NQD

All steps were performed at 4°C. Starting from a 250 ml cell culture, the pellet containing $NQD_{Flag\text{-}H6}$ was resuspended in 30 ml buffer B (20 mM Tris pH 8.0) supplemented with 100 µg/ml lysozyme, 1 µg/ml DNAse I, 30 µl protease inhibitor cocktail (Sigma). After 15 min incubation with agitation, 30 ml buffer C (0.8% Igepal) was added and the cells were incubated with rotation for another 15 min. The cells were disrupted by sonication (using a 750 W sonicator for 5 cycles of 40 s at 50% power output). The lysate was clarified by centrifugation at 30 000g for 30 min at 4°C.

The clarified supernatant was incubated for 1 hr with gentle shaking with 1 ml Chelating Sepharose Fast Flow resin preloaded with Ni^{2+} ions, previously equilibrated in buffer D (10 mM Tris pH 8.0, 0.4% Igepal). The resin was washed with buffer D, and $NQD_{Flag\text{-}H6}$ was eluted in buffer D containing 150 mM imidazole. Eluates were analyzed by SDS-PAGE for the presence of $NQD_{Flag\text{-}H6}$. The fractions containing N were combined, loaded onto a

Superdex 200 HR 10/30 column (Amersham Pharmacia Biotech) and eluted in 10 mM Tris pH 8.0.

Limited proteolysis

Proteolysis was performed at 37°C by dissolving N (2 mg/ml), NQD (0.2 mg/ml), or NI324 (0.2 mg/ml) in 50 mM sodium phosphate buffer pH 7.5, 50 mM NaCl, 0.5 mM DTT. Trypsin was used at an enzyme:substrate ratio varying from 1:1000 to 1:50 (by weight). Proteolysis was stopped after one hour by adding 1 mM PMSF, and digested aliquots were mixed with 2x Laemmli sample buffer and boiled. The extent of proteolysis was assayed by SDS-PAGE and by Western blotting with anti-Flag (Sigma) or anti-Histag (Qiagen) antibodies. Mass analysis of digested fragments was performed as described in (Karlin *et al.*, 2002).

Binding of N and NQD to P

Each co-precipitation assay concerned 2 proteins. A hexahistidine-tagged protein (the precipitating protein) was tested for its ability to precipitate a non-tagged protein (the precipitated protein) on an IMAC resin. All proteins used for co-precipitation studies (except NQD) were prepared according to the same protocol. The induced cells were harvested and divided into 5 ml aliquots. These aliquots were pelleted and frozen at –20°C.

For each assay, two aliquots containing one hexahistidine tagged protein and one untagged protein were resuspended in 500 µl buffer E (10 mM Tris pH 8.0, 10mM imidazole, 0.3% Igepal, 0.5 M NaCl) and sonicated (using a 750 W sonicator and 1 cycle of 6 s at 35% power output). The lysed aliquots were centrifuged 10 min at 13000 g and the supernatants, containing the soluble fractions, were recovered. The soluble fractions of the two aliquots were mixed and immediately incubated with 10 µl TALON resin (Clongech) for 1h at 20°C. The mixture was centrifuged, and both the resin and supernatant were collected. The resin was washed twice with 1 ml buffer E. It was then mixed with 10 µl 2X Laemmli sample buffer, boiled and loaded on a 10% SDS-PAGE gel. For coprecipitation assays involving NQD, a large-scale preparation of $NQD_{Flag\text{-}H6}$ (lysed as described above) was used. 500 µl of a bacterial soluble fraction containing $NQD_{Flag\text{-}H6}$ was mixed to a bacterial lysate expressing the desired protein and to 10 µl TALON resin and tested for coprecipitation as described above.

Immunoprecipitation of N and NQD

Each assay concerned one bacterial lysate (expressing either the wild-type, the mutant nucleoprotein, or none) and one antibody. The antibody was tested for its ability to precipitate the nucleoprotein on a Protein G Sepharose 4FF resin. A culture of either $N_{Flag\text{-}H6}$ or $NQD_{Flag\text{-}H6}$ was induced as described above, the induced cells were harvested and divided into 5 ml aliquots, which were pelleted and frozen at -20°C.

We used either the anti-N mAbs Cl25 or Cl120 described above, or mAbs directed against the Flag (M2, Sigma) or hexahistidine tag (6xHis, Qiagen), herein called aFlag and aH6 respectively .

For each assay, an aliquot containing either the wild-type or mutant N was resuspended in buffer F (10 mM Tris pH 8.0, 0.1% Igepal, 150 mM NaCl) and sonicated (using a 750 W sonicator and 1 cycle of 6 s at 35% power output). All subsequent steps were performed at room temperature. The lysed aliquots were centrifuged 10 minutes at 13000g and the supernatants, containing the soluble fractions, were precleared by incubation for 30 min with 15 µl protein G sepharose. The precleared fractions were mixed with 2 µg mAbs and immediately incubated with 15 µl protein G sepharose for 1h. The mixture was centrifuged, and both the resin and supernatant were collected. The resin was washed twice with 1ml buffer F. It was then mixed with 15 µl 2x Laemmli sample buffer, boiled, and loaded on a 10% SDS-PAGE gel.

Electron Microscopy studies

A drop of a solution of N or NQD was placed on an air glow-discharged carbon-coated grid for a few seconds and then negatively stained with a 2% uranyl acetate aqueous solution. The structures formed by N and NQD were observed in an electron microscope (Philips CM12).

Determination of protein concentration

The theoretical absorption coefficients $\in$ (mg/ml.cm) of N at 280 nm and 260 nm are $\in_{280}$=0.72 and $\in_{260}$=0.48. The theoretical $\in$'s of NQD are the same, since the SL>QD mutation in NQD did not suppress or introduce any tryptophan, phenylalanine or tyrosin. For RNA, $\in_{280}$=11 and $\in_{260}$=20. In the nucleocapsid, each N monomer (58 kDa) binds 6 ribonucleotides. Therefore N represents 96% in mass. Accordingly, the absorption coefficients of the (N:RNA) complex are $\in_{280}$=1.13 (=0.96*0.72+0.04*11) and $\in_{260}$=1.26 (=0.96*0.48+0.04*20), and the corresponding ratio is ($\in_{280}/\in_{260}$)=0.9.

Circular Dichroism

CD spectra of $N_{Flag\text{-}H6}$ and $NQD_{Flag\text{-}H6}$ were recorded on a Jasco 800 dichrograph using 1-mm thick quartz cells in 10 mM sodium phosphate pH 7.0, 0.5 mM EDTA, 0.5 mM PMSF at 20°C. CD spectra were measured between 192 and 260 nm, conditions in which the PM tension was below 600 V. Mean ellipticity values per residue ([Θ]) were calculated as [Θ] = 3300 m ΔA / (l c n), taking into account a path length (l) of 0.1 cm, a number of residues (n) of 539, a molecular mass (m) of 60000 Da, and a protein concentration (c) of 0.21 mg/ml and 0.16 mg/ml for N and NQD respectively, derived from their absorption at 280 nm as described above. Secondary structure contents were estimated using the program CDNN (Dalmas *et al.*, 1994).

Multiple sequence alignment and secondary structure prediction

The multiple alignment of N was generated using the program ClustalW (Thompson et al., 1994) and then manually adjusted with Seaview (Galtier et al., 1996). The secondary structure prediction was carried out using the program PHD (Rost, 1996).

ACKNOWLEDGEMENTS

We wish to thank M.-P. Egloff for help with the design of the mutants, H. Dutartre for help with co-precipitation and immunoprecipitation studies, V. Campanacci for help with CD, D. Gerlier for anti-N antibodies, and J. Lepault for EM studies. This work was supported by the CNRS, and by a grant from the Fondation pour la Recherche Médicale (FRM) to D. Karlin. This study has been carried out with financial support from the Commission of the European Communities, specific RTD programme "Quality of Life and Management of Living Resources", QLK2-CT2001-01225, "Towards the design of new potent antiviral drugs: structure-function analysis of Paramyxoviridae RNA polymerase". It does not necessarily reflect its views and in no way anticipates the Commission's future policy in this area.

REFERENCES

Bankamp, B., Horikami, S. M., Thompson, P. D., Huber, M., Billeter, M., and Moyer, S. A. (1996). Domains of the measles virus N protein required for binding to P protein and self-assembly. *Virology* **216,** 272-7.

Bohn, W., Ciampor, F., Rutter, R., and Mannweiler, K. (1990). Localization of nucleocapsid associated polypeptides in measles virus- infected cells by immunogold labelling after resin embedding. *Arch. Virol.* **114,** 53-64.

Brinkmann, U., Mattes, R. E., and Buckel, P. (1989). High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the dnaY gene product. *Gene* **85,** 109-14.

Brizzard, B. L., Chubet, R. G., and Vizard, D. L. (1994). Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. *Biotechniques* **16,** 730-5.

Buchholz, C. J., Retzler, C., Homann, H. E., and Neubert, W. J. (1994). The carboxy-terminal domain of Sendai virus nucleocapsid protein is involved in complex formation between phosphoprotein and nucleocapsid- like particles. *Virology* **204,** 770-6.

Buchholz, C. J., Spehner, D., Drillien, R., Neubert, W. J., and Homann, H. E. (1993). The conserved N-terminal region of Sendai virus nucleocapsid protein NP is required for nucleocapsid assembly. *J. Virol.* **67,** 5803-12.

Buckland, R., Giraudon, P., and Wild, F. (1989). Expression of measles virus nucleoprotein in *Escherichia coli*: use of deletion mutants to locate the antigenic sites. *J. Gen. Virol.* **70,** 435-41.

Curran, J. (1996). Reexamination of the Sendai virus P protein domains required for RNA synthesis: a possible supplemental role for the P protein. *Virology* **221,** 130-40.

Curran, J. (1998). A role for the Sendai virus P protein trimer in RNA synthesis. *J. Virol.* **72,** 4274-80.

Curran, J., Homann, H., Buchholz, C., Rochat, S., Neubert, W., and Kolakofsky, D. (1993). The hypervariable C-terminal tail of the Sendai paramyxovirus nucleocapsid protein is required for template function but not for RNA encapsidation. *J. Virol.* **67,** 4358-64.

Curran, J., and Kolakofsky, D. (1999). Replication of *paramyxoviruses*. *Adv. Virus Res.* **54,** 403-22.

Curran, J., Marq, J. B., and Kolakofsky, D. (1995). An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J. Virol.* **69,** 849-55.

Dalmas, B., Hunter, G. J., and Bannister, W. H. (1994). Prediction of protein secondary structure from circular dichroism spectra using artificial neural network techniques. *Biochem. Mol. Biol. Int.* **34,** 17-26.

De, B. P., Hoffman, M. A., Choudhary, S., Huntley, C. C., and Banerjee, A. K. (2000). Role of NH_2- and COOH-terminal domains of the P protein of human parainfluenza virus type 3 in transcription and replication. *J. Virol.* **74,** 5886-95.

Deshpande, K. L., and Portner, A. (1984). Structural and functional analysis of Sendai virus nucleocapsid protein NP with monoclonal antibodies. *Virology* **139,** 32-42.

Errington, W., and Emmerson, P. T. (1997). Assembly of recombinant Newcastle disease virus nucleocapsid protein into nucleocapsid-like structures is inhibited by the phosphoprotein. *J. Gen. Virol.* **78,** 2335-9.

Finch, J. T., and Gibbs, A. J. (1970). Observations on the structure of the nucleocapsids of some *paramyxoviruses*. *J. Gen. Virol.* **6,** 141-50.

Fooks, A. R., Stephenson, J. R., Warnes, A., Dowsett, A. B., Rima, B. K., and Wilkinson, G. W. (1993). Measles virus nucleocapsid protein expressed in insect cells assembles into nucleocapsid-like structures. *J. Gen. Virol.* **74,** 1439-44.

Galtier, N., Gouy, M., and Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* **12,** 543-8.

Gill, D. S., Takai, S., Portner, A., and Kingsbury, D. W. (1988). Mapping of antigenic domains of Sendai virus nucleocapsid protein expressed in *Escherichia coli*. *J. Virol.* **62,** 4805-8.

Giraudon, P., Jacquier, M. F., and Wild, T. F. (1988). Antigenic analysis of African measles virus field isolates: identification and localisation of one conserved and two variable epitope sites on the NP protein. *Virus Res.* **10,** 137-52.

Gombart, A. F., Hirano, A., and Wong, T. C. (1993). Conformational maturation of measles virus nucleocapsid protein. *J. Virol.* **67,** 4133-41.

Griffin, D. E. (2001). Measles Virus. *In* "Fields Virology" (B.N. Fields, D.M. Knipe, and P.M. Howley, Eds.), 4th ed., pp 1401-1441 Lippincott-Raven, Philadelphia, PA.

Gubbay, O., Curran, J., and Kolakofsky, D. (2001). Sendai virus genome synthesis and assembly are coupled: a possible mechanism to promote viral RNA polymerase processivity. *J. Gen. Virol.* **82,** 2895-903.

Harty, R. N., and Palese, P. (1995). Measles virus phosphoprotein (P) requires the NH_2- and COOH-terminal domains for interactions with the nucleoprotein (N) but only the COOH terminus for interactions with itself. *J. Gen. Virol.* **76,** 2863-7.

Heggeness, M. H., Scheid, A., and Choppin, P. W. (1980). Conformation of the helical nucleocapsids of *paramyxoviruses* and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration. *Proc. Natl. Acad. Sci. USA* **77,** 2631-5.

Heggeness, M. H., Scheid, A., and Choppin, P. W. (1981). the relationship of conformational changes in the Sendai virus nucleocapsid to proteolytic cleavage of the NP polypeptide. *Virology* **114,** 555-62.

Hirano, A., Wang, A. H., Gombart, A. F., and Wong, T. C. (1992). The matrix proteins of neurovirulent subacute sclerosing panencephalitis virus and its acute measles virus progenitor are functionally different. *Proc. Natl. Acad. Sci. USA* **89,** 8745-9.

Homann, H. E., Willenbrink, W., Buchholz, C. J., and Neubert, W. J. (1991). Sendai virus protein-protein interactions studied by a protein-blotting protein-overlay technique: mapping of domains on NP protein required for binding to P protein. *J. Virol.* **65,** 1304-9.

Horikami, S. M., Curran, J., Kolakofsky, D., and Moyer, S. A. (1992). Complexes of Sendai virus NP-P and P-L proteins are required for defective interfering particle genome replication in vitro. *J. Virol.* **66,** 4901-8.

Horikami, S. M., and Moyer, S. A. (1995). Structure, transcription, and replication of measles virus. *Curr. Top. Microbiol. Immunol.* **191,** 35-50.

Horikami, S. M., Smallwood, S., and Moyer, S. A. (1996). The Sendai virus V protein interacts with the NP protein to regulate viral genome RNA replication. *Virology* **222,** 383-90.

Hubbard, S. J. (1998). The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta* **1382,** 191-206.

Hubbard, S. J., Beynon, R.J. (2001). Proteolysis of native proteins as a structural probe. *In* "Proteolytic Enzymes" (Beynon R.J., Bond J.D., Eds.), 2nd ed., pp 233-264. Oxford University Press, Oxford, NY.

Huber, M., Cattaneo, R., Spielhofer, P., Orvell, C., Norrby, E., Messerli, M., Perriard, J. C., and Billeter, M. A. (1991). Measles virus phosphoprotein retains the nucleocapsid protein in the cytoplasm. *Virology* **185,** 299-308.

Karlin, D., Longhi, S., Receveur, V., and Canard, B. (2002). The measles virus phosphoprotein N-terminal domain (PNT) belongs to the natively unfolded class of proteins. *Virology* **In press**.

Kouznetzoff, A., Buckle, M., and Tordo, N. (1998). Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. *J. Gen. Virol.* **79,** 1005-13.

Lamb, R. A., and Kolakofsky, D. (2001). *Paramyxoviridae* : The Viruses and Their Replication. *In* "Fields Virology" (B.N. Fields, D.M. Knipe, and P.M. Howley, Eds.), 4th ed., pp 1305-1340. Lippincott-Raven, Philadelphia, PA.

Liston, P., Batal, R., DiFlumeri, C., and Briedis, D. J. (1997). Protein interaction domains of the measles virus nucleocapsid protein (NP). *Arch. Virol.* **142,** 305-21.

Liston, P., DiFlumeri, C., and Briedis, D. J. (1995). Protein interactions entered into by the measles virus P, V, and C proteins. *Virus Res.* **38,** 241-59.

Myers, T. M., Pieters, A., and Moyer, S. A. (1997). A highly conserved region of the Sendai virus nucleocapsid protein contributes to the NP-NP binding domain. *Virology* **229,** 322-35.

Myers, T. M., Smallwood, S., and Moyer, S. A. (1999). Identification of nucleocapsid protein residues required for Sendai virus nucleocapsid formation and genome replication. *J. Gen. Virol.* **80,** 1383-91.

Nakai, T., Shand, F. L., and Howatson, A. F. (1969). Development of measles virus in vitro. *Virology* **38,** 50-67.

Radecke, F., Spielhofer, P., Schneider, H., Kaelin, K., Huber, M., Dotsch, C., Christiansen, G., and Billeter, M. A. (1995). Rescue of measles viruses from cloned DNA. *EMBO J.* **14,** 5773-84.

Rose, J. K., and Whitt, M. A. (2001). *Rhabdoviridae*: The Viruses and Their Replication. *In* "Fields Virology" (B.N. Fields, D.M. Knipe, and P.M. Howley, Eds.), 4th ed., pp 1221-1241 Lippincott-Raven, Philadelphia, PA.

Rost, B. (1996). PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol.* **266,** 525-39.

Ryan, K. W., and Portner, A. (1990). Separate domains of Sendai virus P protein are required for binding to viral nucleocapsids. *Virology* **174,** 515-21.

Ryan, K. W., Portner, A., and Murti, K. G. (1993). Antibodies to *paramyxovirus* nucleoproteins define regions important for immunogenicity and nucleocapsid assembly. *Virology* **193,** 376-84.

Schneider, H., Kaelin, K., and Billeter, M. A. (1997). Recombinant measles viruses defective for RNA editing and V protein synthesis are viable in cultured cells. *Virology* **227,** 314-22.

Schoehn, G., Iseni, F., Mavrakis, M., Blondel, D., and Ruigrok, R. W. (2001). Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. *J. Virol.* **75,** 490-8.

Sedlmeier, R., and Neubert, W. J. (1998). The replicative complex of *paramyxoviruses*: structure and function. *Adv. Virus Res.* **50,** 101-39.

Shaji, D., and Shaila, M. S. (1999). Domains of Rinderpest virus phosphoprotein involved in interaction with itself and the nucleocapsid protein. *Virology* **258,** 415-24.

Spehner, D., Drillien, R., and Howley, P. M. (1997). The assembly of the measles virus nucleoprotein into nucleocapsid-like particles is modulated by the phosphoprotein. *Virology* **232,** 260-8.

Spehner, D., Kirn, A., and Drillien, R. (1991). Assembly of nucleocapsidlike structures in animal cells infected with a vaccinia virus recombinant encoding the measles virus nucleoprotein. *J. Virol.* **65,** 6296-300.

Stricker, R., Mottet, G., and Roux, L. (1994). The Sendai virus matrix protein appears to be recruited in the cytoplasm by the viral nucleocapsid to function in viral assembly and budding. *J. Gen. Virol.* **75,** 1031-42.

Suryanarayana, K., Baczko, K., ter Meulen, V., and Wagner, R. R. (1994). Transcription inhibition and other properties of matrix proteins expressed by M genes cloned from measles viruses and diseased human brain tissue. *J. Virol.* **68,** 1532-43.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673-80.

Warnes, A., Fooks, A. R., Dowsett, A. B., Wilkinson, G. W., and Stephenson, J. R. (1995). Expression of the measles virus nucleoprotein gene in *Escherichia coli* and assembly of nucleocapsid-like structures. *Gene* **160,** 173-8.

Du fait qu'il n'a pas une structure modulaire, le cœur de N (N_{CORE}, aa 1-400) se prête mal aux études de délétion. En revanche, nous avons montré que par une (double) mutation ponctuelle, il est possible d'obtenir une nucléoprotéine dont les propriétés de polymérisation sont changées, alors même qu'elle a conservé une structure secondaire et des propriétés antigéniques et biochimiques similaire à la nucléoprotéine originale. Ce variant est cependant toujours agrégé, sans qu'il soit possible de dire dans l'état actuel des choses si s'agit d'une agrégation régulière ou aspécifique. Il est possible qu'une seconde mutation ponctuelle au sein d'une autre région impliquée dans l'auto-association de N supprime totalement la polymérisation.

A notre connaissance, la première purification d'une nucléoprotéine recombinante d'un *Mononegavirales* (VSV) a eu lieu en 1993 (Das and Banerjee, 1993). L'article soulignait que cela permettrait d'entreprendre des études structurales. 9 ans plus tard, les seuls progrès structuraux qui ont été accomplis l'ont été par microscopie électronique sur la forme assemblée de la nucléoprotéine du virus de la rage (Schoehn *et coll.*, 2001). La difficulté d'études cristallographiques sur N est expliquée dans l'introduction de l'article. Il semble bien que nous allons devoir avoir recours à des "astuces" techniques pour résoudre la structure tridimensionnelle de N. En dehors des essais de cristallisation en cours sur NQD, une liste non exhaustive inclut :

- la poursuite de la mutagenèse dirigée de N dans l'espoir d'obtenir un variant qui ne soit plus agrégé.

- l'utilisation à notre avantage d'un des nombreux "artefacts" des systèmes de production recombinante. Ainsi, en baculovirus, RV N se polymérise pour former de courts anneaux (d'une dizaine de nucléoprotéines) autour de petits ARN de transfert qui ont exactement la longueur requise pour cela ; ces anneaux beaucoup plus rigides que les nucléocapsides virale ont permis d'obtenir la première structure à basse résolution de N (Schoehn *et coll.*, 2001).

- la combinaison de cette dernière approche avec l'essai systématique de nucléoprotéines de différents *Paramyxoviridae*. Ainsi la nucléoprotéine d'un *Pneumovirinae* est monomérique dans un système d'expression en cellules eucaryotes (Paul Yeo, communication personnelle).

- la dénaturation, suivie par la renaturation de N sur des ARN d'une longueur prédéfinie, par exemple celle correspondant à un anneau (voir plus haut). Pour la nucléocapside du virus de la rougeole, un tour d'hélice comprend 13 monomères de N, chacun associé à 6 nucléotides. La longueur d'un tel ARN serait donc sans doute 66 (6*11),

72 (6*12), ou 78 (6*13) nucléotides – un anneau étant légèrement plus court que le tour d'hélice correspondant. Une telle dénaturation peut s'effectuer avec du chlorure de guanidium (Baker and Moyer, 1988) ou avec du thiocyanate de sodium.

- la reproduction volontaire de l'artefact décrit dans (Schoehn *et coll.*, 2001), en co-exprimant en bactéries N et un ARN d'une longueur adéquate sous contrôle du promoteur T7. Si la production par la T7 polymérase de larges quantités d'ARN qui ne sont pas traduits n'est pas toxique pour la cellule (voir (Miroux and Walker, 1996)), on peut espérer obtenir des anneaux homogènes qui conviendraient pour la cristallisation.

- la construction de nucléocapsides homogènes d'une taille convenant à la cristallisation, à partir de virus défectifs interférents de longueur minimale (par le groupe de L. Roux)

En revanche, la fusion de protéines (telles que la β-galactosidase) à N dans l'espoir d'empêcher sa polymérisation semble ne pas marcher : les chimères résultantes sont instables (Warnes *et coll.*, 1995).

Il y a aussi la possibilité d'étudier une autre conformation de N, complexée à P : N°-P, une approche suivie au laboratoire. R. Ruigrok et D. Blondel disposent déjà du complexe N°-P du RV (Schoehn *et coll.*, 2001). Notons qu'à première vue, Green *et coll.* disposent de cristaux du complexe N°-P du VSV (Green *et coll.*, 2000). Pourtant, il semble que la co-expression de N° et P ait un rôle différent entre VSV et d'autres virus tels que RV (un *Rhabdoviridae* aussi, pourtant) ou les *Paramyxoviridae*. En effet, N de VSV est difficilement soluble en l'absence de P. Il est possible que le complexe N-P de VSV soit en fait un complexe N^{NUC}-P …

Il serait intéressant d'établir si NQD possède toujours **l'épitope conformationnel** reconnu par l'anticorps conformationnel N46 - dont nous disposons au laboratoire - (Gombart *et coll.*, 1993), formé par la région 23-239 de N. Celle-ci comprend le site de la double substitution de NQD (aa 228-229). De plus, l'anticorps N46 bloque l'interaction N°-P, ce qui implique que la région 23-239 de N° joue un rôle dans la liaison à P. Il sera tout aussi intéressant d'étudier la liaison de NQD à P en *co-expression*, pour avoir éventuellement accès à un **complexe (NQD°-P)**. Enfin, peut-être NQD pourrait-il apporter des précisions sur le site de **liaison à l'ARN** de N. En effet, celui-ci est normalement caché à l'intérieur de la nucléocapside et n'est sans doute pas accessible à des expériences de pontage ("cross-linking"), ce qui reste à démontrer. Il y a peut-être une chance que ce site soit exposé dans NQD - le fait que NQD ne contienne pas d'ARN après purification n'exclut pas qu'il contienne un site de liaison à l'ARN.

Par ailleurs, notre article, couplé à l'article originel (Bankamp *et coll.*, 1996), indique que la région 189-239 de N est importante pour l'auto-association de N, et renforce l'idée que le domaine de liaison à P (N_{TAIL}) est indépendant du reste de N, puisque les deux variants de N qui ne s'assemblent plus de façon régulière continuent à se lier à P. Cette hypothèse est explicitée dans la revue, puisque N_{TAIL} contient presque certainement des régions non structurées. Par conséquent, nous allons produire NQD sous une forme tronquée (sans sa queue C-terminale) en vue de sa cristallisation. Cet article illustre d'une part la faisabilité de l'approche de mutagenèse dirigée des régions d'auto-association de N (même si notre intention n'était pas d'obtenir des agrégats homogènes, mais d'empêcher la polymérisation de N), et d'autre part la dépense importante en temps et en effort de purification qu'elle requiert. Si une telle approche devait être étendue à d'autres mutants, un abord plus systématique serait nécessaire (puisque seul un mutant sur quatre est intéressant pour la cristallisation) et envisageable seulement avec un accès à des outils de mutagenèse, d'expression et de purification à haut débit.

Un aspect en apparence mineur de l'article, mais qui pourrait avoir de grandes conséquences pratiques, est la possibilité d'introduire un épitope artificiel en fusion N-terminale à N. Cela pourrait être d'une grande aide dans la caractérisation des interactions de N à l'aide d'anticorps. Enfin, ce travail a permis de mettre au point une méthode de purification du complexe N^{NUC}-P qui constitue un prélude à leur étude détaillée, qui se poursuit au laboratoire.

REVUE : "STRUCTURAL FLEXIBILITY WITHIN THE REPLICATIVE COMPLEX OF *PARAMYXOVIRINAE*"

RESUME

Flexibilité structurale au sein du complexe réplicatif des *Paramyxovirinae*

Nous passons en revue les progrès récents dans la caractérisation du complexe réplicatif des *Paramyxovirinae* d'un point de vue structural et fonctionnel. En particulier, nous soulignons la flexibilité structurale de la phosphoprotéine virale (P) et de la nucléoprotéine (N). Les régions non structurées de N et de P sont impliquées dans des interactions avec plusieurs partenaires protéiques qui sont critiques pour la réplication. En particulier, nous proposons que la conformation étendue de ces régions puisse faciliter la formation d'un complexe tripartite entre la forme non assemblée de N, P, et la polymérase virale. Une certaine flexibilité pourrait aussi être requise pour l'interaction entre N et la protéine de matrice (M) durant l'assemblage du virus. Nous montrons que la flexibilité structurale est probablement une propriété répandue chez les virus à ARN simple brin non segmenté. Ces virus constituent un système modèle fascinant pour l'étude des rôles fonctionnels des régions désordonnées des protéines.

INTRODUCTION

Après la découverte du caractère intrinsèquement désordonné de la région PNT des phosphoprotéines des *Paramyxovirinae* (voir article n°1), il était intéressant de poursuivre plus en détail l'étude du désordre structural au sein du complexe réplicatif, et de ses implications possibles. Ainsi, des expériences de protéolyse limitée sur N suggèrent très fortement que sa queue C-terminale (N_{TAIL}) contient des régions désordonnées. Pourtant, à notre connaissance, aucun article de recherche ne mentionne cette hypothèse.

Nous nous sommes appuyés sur les progrès récents en matière de cartographie fonctionnelle des protéines N et P, (en particulier chez les *Rubulavirus*) ainsi que sur l'élucidation de la structure tridimensionnelle de plusieurs régions de P de SeV. Dans cette revue, nous avons mis en lumière l'importance des régions désordonnées des protéines N et P chez tous les *Paramyxovirinae* (aussi bien par leur simple abondance que d'un point de vue fonctionnel). Ce travail devrait ainsi éviter de nombreuses tentatives (vouées à l'échec) de cristallisation des régions non structurées que nous avons décrites.

Notre étude confirme à quel point l'organisation de P est modulaire. En particulier, nous avons montré qu'une région de P qui n'avait jamais été caractérisée a de fortes chances

d'être intrinsèquement désordonnée. Nous avons aussi mis en évidence le fait que les protéines N et P des 3 genres des *Paramyxovirinae* présentent des caractères fonctionnels conservés, mais que les régions associées à ces caractères (qui sont toujours désordonnées) peuvent se situer à des endroits différents de N et de P, suivant les genres. Cette organisation de N et de P, jamais soulignée auparavant, a des implications immédiates en termes d'association entre elles et avec L. En nous appuyant sur des résultats très récents, nous proposons un modèle d'association entre N, P, et L durant la réplication.

En filigrane, nous avons étendu ce travail à la famille des *Rhabdoviridae*, proche des *Paramyxoviridae*, dont les protéines P et N contiennent très probablement des régions non structurées et/ou flexibles. Par contre, il sera intéressant d'examiner en détail les *Pneumovirinae*, chez qui N et P sont de poids moléculaires beaucoup moins élevés, et qui à première vue pourraient être majoritairement ordonnées. Une telle observation pourrait nous éclairer sur le rôle du désordre structural dans le complexe réplicatif.

Par ailleurs, cette étude nous a conduit à émettre l'hypothèse que des protéines spécifiées par des cadres de lecture chevauchants (comme c'est le cas de P, C, et V) contiennent des régions désordonnées avec une fréquence supérieure à la moyenne. La vérification de cette hypothèse fait l'objet d'une collaboration avec K. Dunker.

Enfin, notre étude soulève des problématiques nouvelles, pour lesquelles nous proposons des pistes de recherche : quel est le rôle du désordre structural et de la flexibilité au sein du complexe réplicatif ? et durant l'assemblage du virus ? Les régions désordonnées des protéines se structurent-elles en se liant à leur(s) partenaire(s) ? Est-ce une coïncidence si d'autres protéines virales contiennent des régions (sans doute désordonnées) qui sont variables en séquence et hautement phosphorylées ?

Nous espérons que ce travail inspirera d'autres études et que le complexe réplicatif des *Mononegavirales* deviendra un système modèle dans l'étude du désordre structural.

Structural flexibility
in the replicative complex of *Paramyxovirinae*

David Karlin, Sonia Longhi, and Bruno Canard*

Running title: Structural flexibility in *Paramyxovirinae*

Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Université Aix-Marseille I et II, ESIL, Campus de Luminy, 13288 Marseille Cedex 09, France

(*) to whom requests for reprints should be addressed
ESIL-CNRS-AFMB Case 925
163, avenue de Luminy
13288 Marseille Cedex 09
France
E-mail: **bruno@esil.univ-mrs.fr**
Tel: (33) (491) 82 86 44
Fax: (33) (491) 82 86 46

Soumis à *Journal of Virology*

ABBREVIATIONS

bPIV-3: bovine parainfluenza virus type 3; CDV: canine distemper virus; EM: electron microscopy; hPIV-2: human parainfluenza virus type 2; hPIV-3: human parainfluenza virus type 3; L: large subunit of the polymerase; LB: L-binding site; LDR: long disordered region (>40 aa); MV: measles virus; MuV: mumps virus; N: nucleoprotein; (n): neosynthesized nucleoprotein; N°: soluble form of the nucleoprotein; N^{NUC}: assembled form of the nucleoprotein; P: phosphoprotein; PCT: phosphoprotein C-terminal moiety; PMD: phosphoprotein multimerization domain; PNT: phosphoprotein N-terminal moiety; PONDR: predictor of disordered regions; PPRV: peste-des-petits-ruminants virus; PX: protein X; RNP: ribonucleoprotein complex; RV: rabies virus; SeV: Sendai virus; SV5: simian virus type 5; VSV: vesicular stomatitis virus; XD: C-terminal domain of the protein X.

ABSTRACT

We review recent advances in characterizing the replicative complex of *Paramyxovirinae* from a structural and functional point of view. In particular, we emphasize the structural flexibility encountered in the viral phosphoprotein (P) and in the nucleoprotein (N). Unstructured regions of N and P are involved in interactions with several protein partners, which are critical for replication. In particular, we propose that the extended conformation of these regions might facilitate the formation of a tripartite complex between the unassembled form of N, P, and the viral polymerase. Flexibility might also be required for proper interaction of N with the matrix protein (M) during virus assembly. We show that structural flexibility is likely to be a widespread property encountered within non-segmented, negative-strand RNA viruses. These viruses provide a fascinating model system for studying the functional roles of disordered regions in proteins.

INTRODUCTION

The family *Paramyxoviridae*, which includes major human pathogens such as measles virus, is composed of enveloped viruses with a non-segmented, negative, single-stranded RNA genome. It is subdivided into two subfamilies: *Paramyxovirinae* and *Pneumovirinae*, distinguished by morphological as well as functional features. *Paramyxovirinae* are composed of three main genera: *Morbilliviruses* (prototype: measles virus), *Respiroviruses*, formerly called *Paramyxoviruses* (prototype: Sendai virus), and *Rubulaviruses* (prototype: human parainfluenza virus type 2). In recent years, emergent viruses have been discovered and will likely be placed into new genera, such as the genus *Henipavirus* proposed for Nipah and Hendra viruses (Wang *et al.*, 2000). Likewise, it is considered to place Newcastle disease virus (NDV), presently a *Rubulavirus*, in a new genus. We will center this study around the three best-characterized genera, calling in other viruses when they present informative differences.

VIRION STRUCTURE

Paramyxovirinae virions are enveloped by a lipid bilayer derived from the plasma membrane of the host cells. Two viral glycoprotein spikes are inserted into the membrane: the attachment protein (H), and the fusion protein (F). At the inner surface of the virus envelope lies the matrix protein (M). Genomic RNA is packaged by the nucleoprotein (N) within a helical nucleocapsid. Transcription and replication are carried out on this (N:RNA) template by the viral RNA-dependent RNA polymerase, formed by a complex between the phosphoprotein (P) and the large protein (L). These three proteins form the viral ribonucleoprotein complex (RNP) (figure 1). Incorporation of viral RNPs during virus assembly is ensured by an interaction between N and M, which is thought to be the central organizer of virus morphogenesis, possibly through other contacts with F and H. Thus, N, P and L are involved in manifold protein-protein and protein-RNA interactions. Two other accessory viral proteins, C and V, modulate transcription and replication. They also play a role in pathogenicity, virulence, and apoptosis. For a detailed description of the virion structure and the viral cycle, see (Lamb and Kolakofsky, 2001).

Figure 1

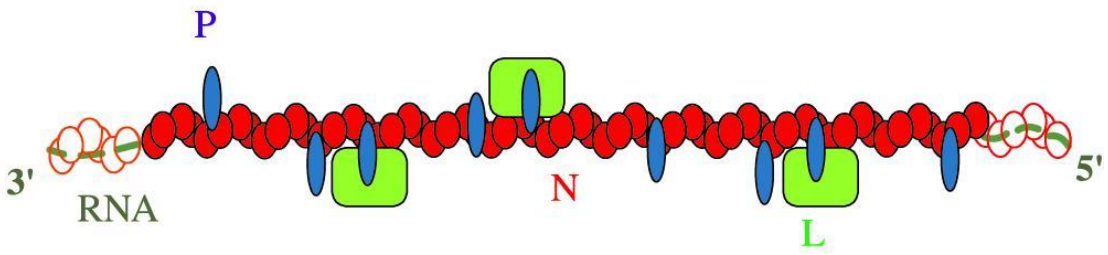

Figure 1. Viral ribonucleoprotein complex (RNP)

The components of the RNP are depicted as: red ovals, nucleoprotein (N); green rectangle, large polymerase subunit (L); blue ovals, tetramers of the phosphoprotein (P); green dotted line, encapsidated genome RNA. N packages RNA in a helical nucleocapsid, represented symbolically here. The 3' ends and 5' ends of encapsidated RNA, are visible through the nucleoproteins at both ends of the nucleocapsid. Note that they do not protrude from the nucleocapsid. Binding of one molecule of L to the nucleocapsid is mediated by one P tetramer. In actively transcribing RNPs, P is found in clusters around L, but supplementary P molecules are also found away from L, a fact poorly understood.

REPLICATIVE CYCLE

As their name implies, negative-strand RNA viruses have a genome of polarity opposed to that of viral mRNAs. Once the viral RNPs are released in the cytoplasm of infected cells, the viral polymerase first transcribes the genome sequentially into mRNAs, which are in turn translated to yield viral proteins. In particular, N and P accumulate in sufficient number to sustain replication. Upon an unknown signal, the polymerase switches into replicative mode, and synthesizes a full-length RNA antigenome, which is concurrently encapsidated by N. The antigenome serves in turn as a template for replication of genomes (figure 2). For a detailed description of the replicative cycle, see (Curran and Kolakofsky, 1999; Lamb and Kolakofsky, 2001; Sedlmeier and Neubert, 1998).

At least two forms of N exist in infected cells: a soluble form, N°, which is used for encapsidating RNA, whereas the form of N assembled onto RNA is called N^{NUC}. In *Respiroviruses* and *Morbilliviruses*, N has the capacity to self-assemble on cellular RNA in the absence of any other viral protein to form nucleocapsid-like particles. Therefore, a mechanism for preventing illegitimate self-assembly of N is necessary. This role is played by P, which forms a soluble N°-P complex, which is used by the polymerase as the substrate for encapsidation. This does not preclude other unknown factors that would regulate the assembly of N. In particular, the role of cellular proteins in replication is discussed in (Choudhary *et al.*, 2000; De and Banerjee, 1999; De *et al.*, 1997; Lai, 1998).

Although the roles of N, P, and L within the replicative complex have been partially clarified, at present we do not have by far a detailed picture of RNA synthesis. In particular, very limited detailed three-dimensional information on the replicative machinery of *Paramyxovirinae* is available. This situation stems from several facts: the difficulty to obtain homogenous polymers of N suitable for X-ray analysis (Karlin *et al.*, 2002a; Schoehn *et al.*, 2001); aggregation of recombinant viral proteins ((Gupta and Banerjee, 1997), and our unpublished observations); the very large size of L, impeding the production of suitable quantities; the structural flexibility of N and P. We will show that the latter factor is probably underestimated. At present, most structural information has been obtained using a combination of electron microscopy (EM), deletion analysis, limited proteolysis studies, and epitope mapping.

This review aims at giving an overview of the replicative machinery of *Paramyxovirinae*, with an emphasis on the role of structural flexibility. We will discuss structural data in a functional context, presenting common themes and pointing out the most informative differences. Whenever possible, illustrated models are presented. To avoid cluttering the text, precise location of functional regions are given in the corresponding figures.

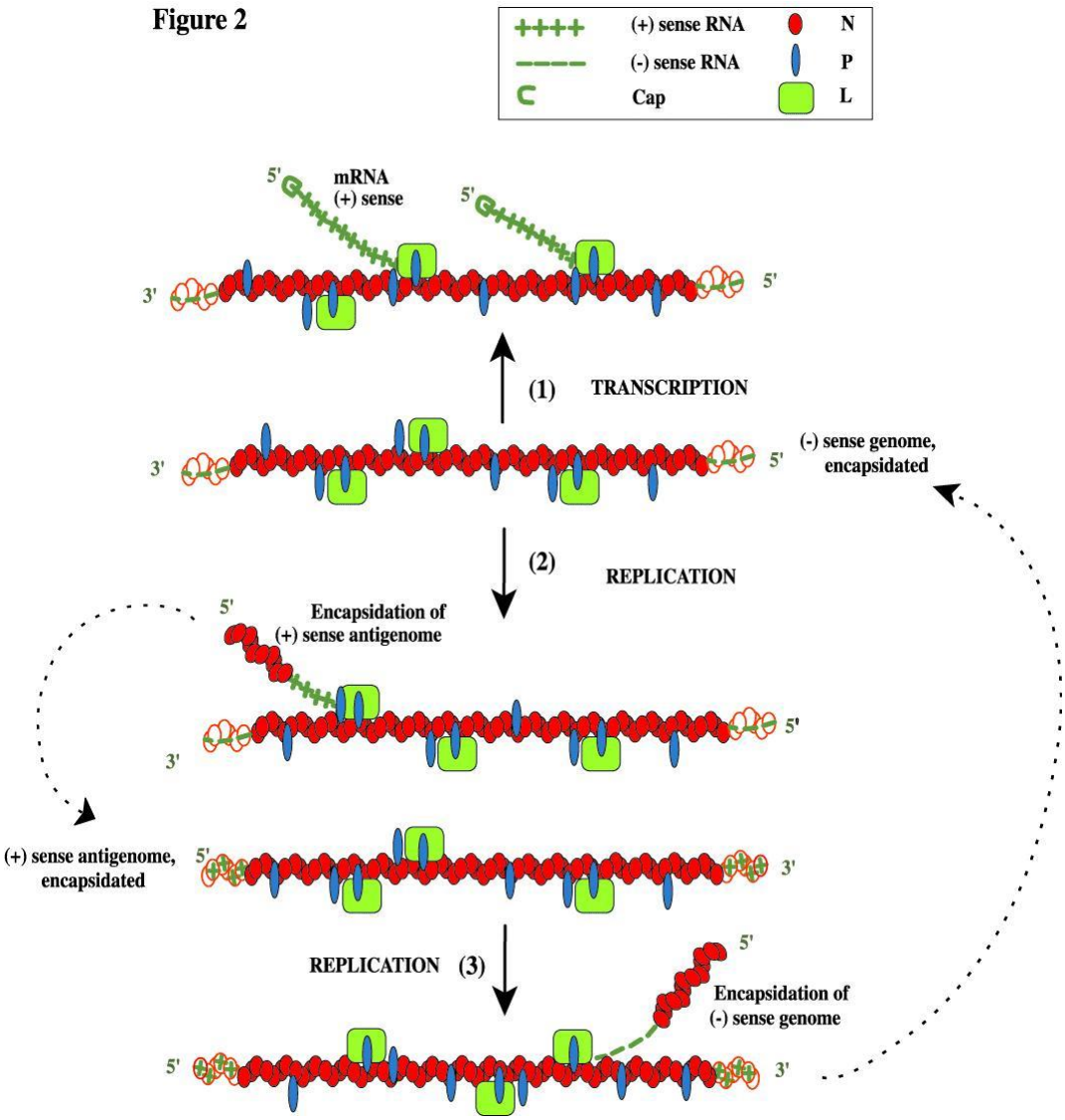

Figure 2. Replicative cycle of the negative-stranded RNA viruses *Paramyxoviridae*

The conventions are the same as in Figure 1. Genomic RNA is defined as being of (-) polarity, and viral mRNA of (+) polarity. (+) RNAs are represented as stretches of green (+) signs, whereas (-) RNAs are represented as stretches of green (-) signs.

1) Genomic RNA is being transcribed into mRNA. Note that the transcribed mRNAs are capped, as symbolized by a half-circle at the 5' end of the mRNA.

2) Genomic RNA is being replicated into a full-length complementary copy, the anti genome, of (+) polarity. The antigenome is encapsidated concurrently to its synthesis; it is not known, however, whether the nascent antigenome is encapsidated as soon as synthesized, or with a slight delay, as depicted here for clarity.

3) Antigenomic RNA is in turn being replicated into a genomic RNA, which is concurrently encapsidated. This new genome RNA is a template for further mRNA synthesis, called secondary transcription (not shown).

MAIN ACTORS OF THE REPLICATIVE MACHINERY

THE NUCLEOPROTEIN (N)

The nucleoprotein (524 to 545 aa) is the main structural protein of the viral ribonucleoprotein complex, in which it plays an important functional role:

- N interacts with P and with the P-L complex during **transcription**. It is not known whether N plays a passive or active role in presenting RNA to the polymerase.
- N might play a role in the **switch between transcription and replication**.
- N selectively **encapsidates** viral RNA during replication (this process is probably coordinated by the polymerase). This function itself might be divided in: i) specific binding to a sequence located at the 5' end of genomes and antigenomes (called trailer and leader sequence, respectively) and ii) non-specific RNA binding.
- N interacts with the matrix protein (M) during viral **assembly**.

The organization of N is detailed in figure 3. Deletion analysis and Electron Microscopy (EM) studies have shown that *Paramyxoviridae* N are divided in two regions structurally and functionally distinct: an N-terminal moiety, N_{CORE}, well conserved in sequence, and a hypervariable, C-terminal moiety, N_{TAIL}.

Figure 5

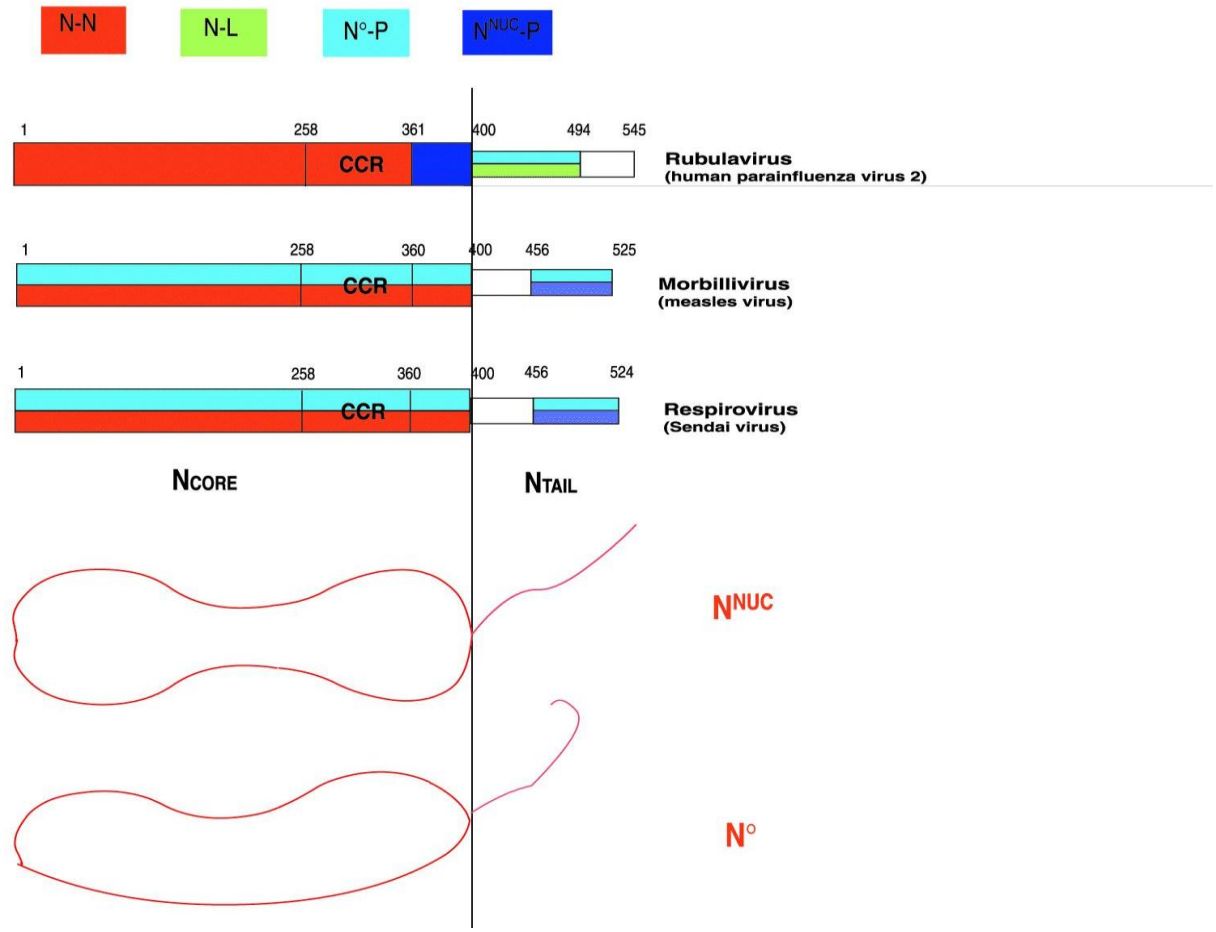

Figure 3. Comparison of the organization of N in *Rubulaviruses, Morbilliviruses,* and, *Respiroviruses*.

The structural and functional organization of N in the 3 genera is presented in the upper panel. A schematic picture of the corresponding proteins is found in the lower panel.

Upper panel

N_{CORE} : N-terminal moiety of N. N_{TAIL}, C-terminal moiety. CCR: central conserved region. Globular regions are represented as large boxes, and unstructured or flexible regions as narrow boxes. The regions of interaction between N and itself, or P and L, are identified by a color code (above the diagram). The precise regions of N-N and N°-P interaction are not known with certainty.

Lower panel

Globular regions are contoured, whereas unstructured regions are represented by lines. Although no high-resolution structural data are available for *Paramyxoviridae* N, it is known that there are several regions of self-interaction in N (see text), like in rabies virus (RV) N (Schoehn *et al.*, 2001). An individual N^{NUC} molecule has been depicted with a bilobal form similar to that of RV N solely to symbolize this fact. N°, which shares a conformational epitope with N^{NUC} (Gombart *et al.*, 1993), has been represented with a different conformation but a conserved half.

N_{TAIL}

Within nucleocapsids, N_{TAIL} protrudes from the globular body of N_{CORE} (Heggeness *et al.*, 1980; Heggeness *et al.*, 1981). N_{TAIL} is required for stable binding of N° to P (Curran *et al.*, 1993; Harty and Palese, 1995; Nishio *et al.*, 1999). *Morbillivirus* and *Respirovirus* N_{TAIL} contain the region responsible for binding of nucleocapsids to P (Bankamp *et al.*, 1996; Buchholz *et al.*, 1994; Liston *et al.*, 1997; Ryan *et al.*, 1993) and consequently N_{TAIL} is required for N to act as a template for viral RNA synthesis (Buchholz *et al.*, 1994; Curran *et al.*, 1993). *Rubulavirus* N_{TAIL} contains an L-binding site that is not found in *Morbillivirus* and *Respiroviruses* (Nishio *et al.*, 2000). The M-binding region of nucleocapsids has been recently mapped to N_{TAIL} in *Respiroviruses* (Coronel *et al.*, 2001). This was expected considering the accessibility of N_{TAIL} on the surface of the nucleocapsid, and because N_{TAIL} is negatively charged, whereas M is highly positively charged. However, no comparable data are available for other genera.

The fact that N_{TAIL} is not seen in EM and is hypersensitive to proteolysis (Deshpande and Portner, 1984; Heggeness *et al.*, 1981; Karlin *et al.*, 2002a) strongly suggests that it is mostly unstructured within nucleocapsids, at least after isolation (Fontana *et al.*, 1997; Hubbard, 1998; Hubbard, 2001). There is a growing awareness that numerous proteins or protein domains are unstructured under physiological conditions, in the absence of a binding partner. These proteins are said to be intrinsically disordered, or natively unfolded (Romero *et al.*, 1997; Weinreb *et al.*, 1996). In recent years, it has been discovered that these proteins or protein domains are usually distinguished from globular proteins by common features of sequence composition. Intrinsically disordered proteins tend to have a low sequence complexity (i.e. they make use of fewer types of amino acids) (Romero *et al.*, 1999; Romero *et al.*, 2001). They are also enriched in most amino acids that are preferred at the surface of globular proteins (A, R, G, Q, S, P, E and K) (termed "disorder-promoting amino acids") and depleted in W, C, F, I, Y, V, L and N ("order-promoting amino acids") (Williams *et al.*, 2001). A neural network that can predict regions of disorder, PONDR, was successfully developed by the group of Dunker and Obradovic (Li *et al.*, 1999; Romero *et al.*, 1997). A borderline length of 40 residues has been chosen in an extensive study to define Long Disordered Regions (LDRs) of proteins with very good confidence (Romero *et al.*, 1998). A second method developed by Uversky *et al.* relies on the net charge/hydrophobicity ratio of proteins, which is distinctly higher for fully disordered proteins than for their structured counterparts. This allows to discriminate between the two classes with a very good accuracy (Uversky, 2002; Uversky *et al.*, 2000). However, contrary to PONDR, this method can only be applied to modular regions, and therefore requires prior knowledge of the organization of the protein under study. We have investigated the sequence properties of *Paramyxovirinae* N_{TAIL} to determine whether its non-globular conformation might be an intrinsic property. PONDR

predicts a LDR in N_{TAIL} in the 3 genera (figure 4A). Furthermore, the sequence properties of N_{TAIL} (namely low hydrophobicity and relatively high net charge due to the presence of numerous acidic residues (Uversky *et al.*, 2000)) are typical of an intrinsically disordered protein (figure 4B). Thus, the sequence properties of N_{TAIL}, coupled to its hypersensitivity to proteolysis, converge to the idea that N_{TAIL} is mostly unstructured in the absence of a binding partner. The great sequence variability of N_{TAIL} provides a further indication that sequence hypervariability might be an indicator of structural disorder (Brown *et al.*, 2002).

Figure 4

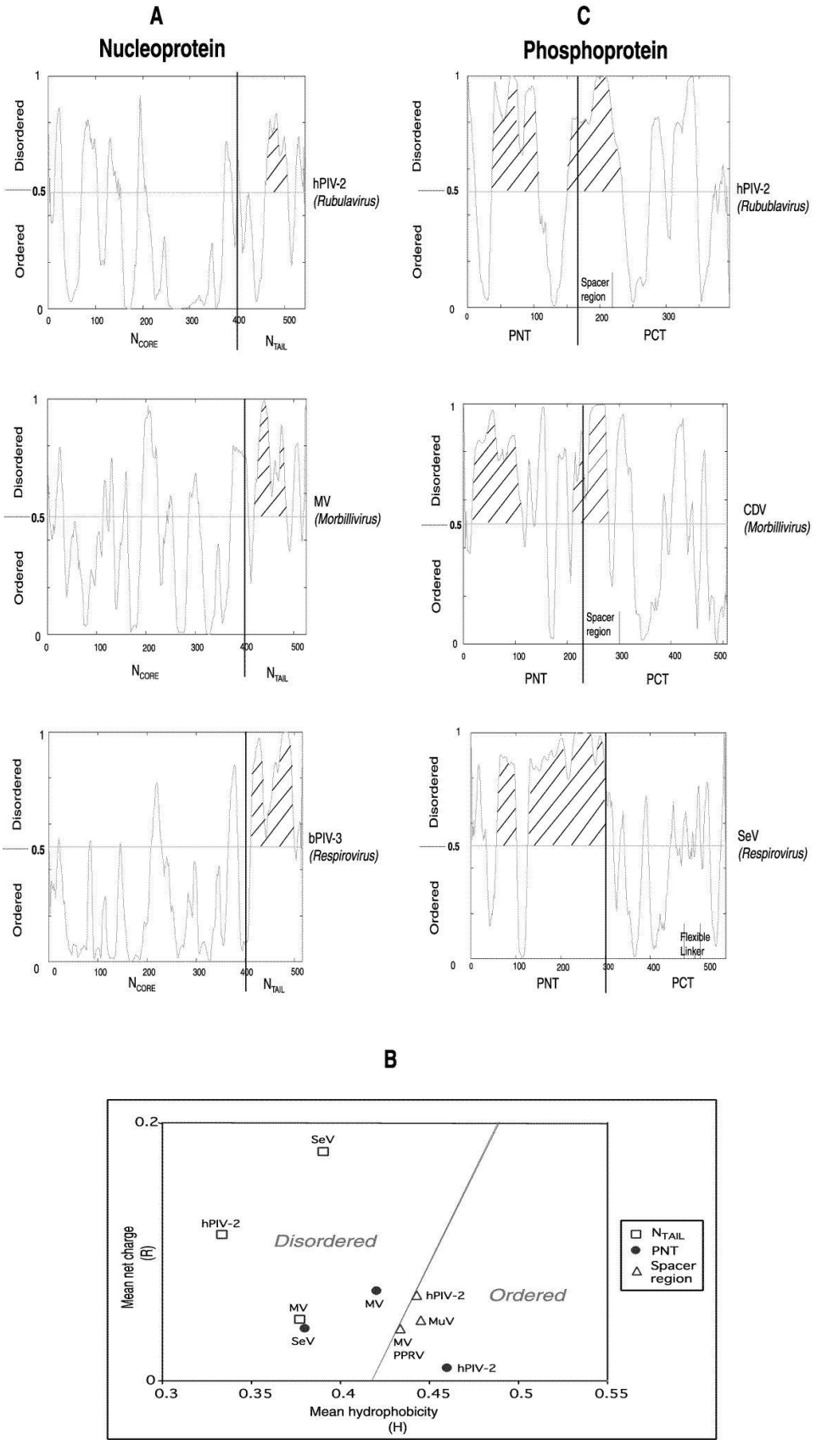
A
Nucleoprotein
C
Phosphoprotein
Disordered
Ordered
0.5
hPIV-2
(Rubulavirus)
MV
(Morbillivirus)
bPIV-3
(Respirovirus)
CDV
(Morbillivirus)
SeV
(Respirovirus)
N_CORE
N_TAIL
PNT
PCT
Spacer region
Flexible Linker
B
Mean net charge (R)
Mean hydrophobicity (H)
Disordered
Ordered
SeV
hPIV-2
MV
MuV
PPRV
N_TAIL
PNT
Spacer region

Figure 4. Predictions of structural disorder in different regions of N and P

A. PONDR Prediction for nucleoproteins of each Paramyxovirinae genus. Disorder prediction values for a given residue are plotted against the residue number. The significance threshold, above which residues are considered to be disordered, set to 0.5, is shown. Long (> 40 residues) Disordered Regions (LDRs) are hatched. The bold, vertical line separates the N and C-terminal moieties of N, N_{CORE} and N_{TAIL}. In each genus, PONDR predicts one LDR in N_{TAIL} but none in N_{CORE}.

B. Net charge / hydrophobicity plot of different regions of N and P

The mean net charge (R) of a protein region is plotted against its mean hydrophobicity (H), calculated as in (Karlin et al., 2002b). The charge/hydrophobicity diagram is divided into two regions by a line, which corresponds to the equation H = (R+1.151)/2.785. In the left of the diagram (where H < (R+1.151)/2.785), a protein is predicted to be disordered, whereas it is predicted to be structured in the right part. It can be seen that hPIV-2 PNT is in the globular area; so are the spacer regions of all viruses, although they are borderline. All N_{TAIL} regions are in the intrinsically disordered area.

C. PONDR prediction for phosphoproteins of each Paramyxovirinae genus.

The conventions are the same than in A. The bold vertical line separates the N and C-terminal modules of P, PNT and PCT.

The predictions are qualitatively similar for other members of the genera (same number of LDR and approximately same position) (not shown). The spacer region of Rubulaviruses and Morbilliviruses is predicted to be disordered, whereas there is no comparable region in Respiroviruses, in agreement with (Tarbouriech et al., 2000). PONDR does not assign a LDR to the flexible linker of SeV (aa 474-517). However, note the pattern of predicted disorder alternating with borderline order at this position. Likewise, it cannot be excluded that there is such a flexible region at the corresponding location of Morbillivirus and Rubulavirus P, according to their PONDR prediction.

Sequences for this study were obtained using the National Center for Biotechnology Information's (NCBI) sequence retrieval software. P and N sequence accession numbers are as follow: bPIV-3 N: AAF28254; MV N: P35972; SeV N: Q07097; hPIV-2 N: P21737; MV P: CAA91364; hPIV-2 P: P23056; MuV P: P16072; PPRV P: Q91QS4; SeV P: P04860.

N_{CORE}

Respirovirus and *Morbillivirus* N_{CORE} contain all regions necessary for self-assembly and RNA-binding, since nucleoproteins composed only of the core region can encapsidate neosynthesized RNA into nucleocapsid-like particles (Buchholz *et al.*, 1993; Curran *et al.*, 1993). SeV N_{CORE} contains a region required for the nucleocapsid to function as a template in viral genome replication (aa 114-129), perhaps by mediating a change in conformation (Myers and Moyer, 1997). Thus, the template function of N is not restricted to merely interacting with P and (P-L) through its tail, rather N may have an active contribution to replication. Within N_{CORE}, deletion studies have failed to identify independent, modular domains, but have identified regions involved in N-N interaction. The region spanning aa 258-360, called Central Conserved Region (CCR) (figure 3) , is well conserved in sequence and mainly hydrophobic. The CCR constitutes one of the regions of self-association (Liston *et al.*, 1997; Myers *et al.*, 1997; Myers *et al.*, 1999) and is also involved in RNA binding (Kouznetzoff *et al.*, 1998; Myers *et al.*, 1997; Myers *et al.*, 1999). Another region important for self-association is contained within residues 189-239 of MV N (Bankamp *et al.*, 1996; Karlin *et al.*, 2002a).

Several features distinguish *Rubulavirus* N proteins from those of other *Paramyxovirinae*. In the three genera, large cytoplasmic inclusion bodies that contain nucleocapsids associated with P can be visualized during infection (Bohn *et al.*, 1990; Lamb and Choppin, 1977; Precious *et al.*, 1995). *Respirovirus* and *Morbillivirus* N forms such inclusion bodies when expressed alone in mammalian cells, but *Rubulavirus* do not. Rather, inclusion bodies are formed only upon co-expression of P (Precious *et al.*, 1995). Whether this means that N does not form nucleocapsid-like particles in the absence of P is unknown and would require EM studies. In agreement with these different morphological features, the *Rubulavirus* N^{NUC}-P binding site is located in N_{CORE} (aa 295-400) (Nishio *et al.*, 1999). Nucleoproteins containing at least the first 295 aa form aggregates, whereas any deletion within this region abolishes self-interaction (Nishio *et al.*, 1999). However, the aggregates formed by aa 1-295 of N were not characterized. Given the high sequence conservation of N_{CORE}, and in particular, that of aa 296-360 throughout *Paramyxovirinae*, it is likely that they form unstructured agregates, as those formed by deletion mutants of SeV and MV N_{CORE} (Bankamp *et al.*, 1996; Buchholz *et al.*, 1993). Therefore, we have represented aa 1-360 as being necessary for *Rubulavirus* N-N interaction in figure 3. Likewise, the N^{NUC}-P binding site has been represented as aa 361-400.

The location of the RNA-binding site(s) within N_{CORE} is unknown. When denatured, such as in Northwestern blots, N does not bind RNA indicating that the RNA-binding site is probably formed by maturation of N during encapsidation (Lamb and Kolakofsky, 2001). In agreement, all mutants in which self-association is impaired do not package RNA (Karlin *et*

al., 2002a; Myers *et al.*, 1997; Myers *et al.*, 1999). More subtle mutations in the region 360-375 were found that did neither disrupt RNA-binding nor the morphology of N, but rendered N inactive in replication (Myers *et al.*, 1999). The author suggested that particular residues in this region might be involved in binding the leader sequence of SeV RNA, versus non-specific RNA.

THE PHOSPHOPROTEIN (P)

The phosphoprotein (named so because it is highly phosphorylated) plays a central role in both transcription and replication.

Like N, P performs a number of very different functions :

- P is an indispensable **subunit of the viral polymerase** complex, needed for the stability of L. It mediates **L-binding** to the (N:RNA) template. Whether it has catalytic activity of its own is unknown.
- P performs a supplementary function involving **binding of additional P molecules** to the (N:RNA) template, independently of its complex formation with L.
- P **prevents illegitimate self-assembly of N** by forming an (N°-P) complex.
- During replication, P allows the polymerase to **deliver N to the nascent RNA**.

From a structural point of view, P is the best-characterized protein of the replicative complex. The general organization of P is conserved within *Paramyxovirinae*, though some features differ between the three genera. P is a modular protein, organized into two moieties functionally and structurally distinct: the N-terminal moiety (PNT), and the C-terminal moiety (PCT). One most interesting feature of the P gene is that the PNT domain is expressed either fused to a zinc-binding region (protein V), and/or as a protein of its own (called protein W), which has greatly helped to clarify the role of PNT. Indeed, all *Paramyxovirinae* P genes encode at least two proteins using overlapping reading frames. The most spectacular example is the SeV P gene that yields eight primary translation products by using a combination of ribosomal choice and cotranscriptional mRNA editing: W, C, C', Y1, Y2, P, D, and PX (reviewed in (Lamb and Kolakofsky, 2001; Latorre *et al.*, 1998)). In figure 5, only the proteins that share a part of their sequence with P are represented (namely P, V, W, and the protein X (PX)).

Most functional data concerning P come from studies on SeV. SeV PCT, which is the most conserved in sequence, contains all the regions required for transcription (Curran, 1996). On the other hand, the poorly conserved PNT provides several additional functions required for replication (Curran, 1996; Curran *et al.*, 1995b; Curran *et al.*, 1994). The organization of P is depicted in figure 6.

Figure 5

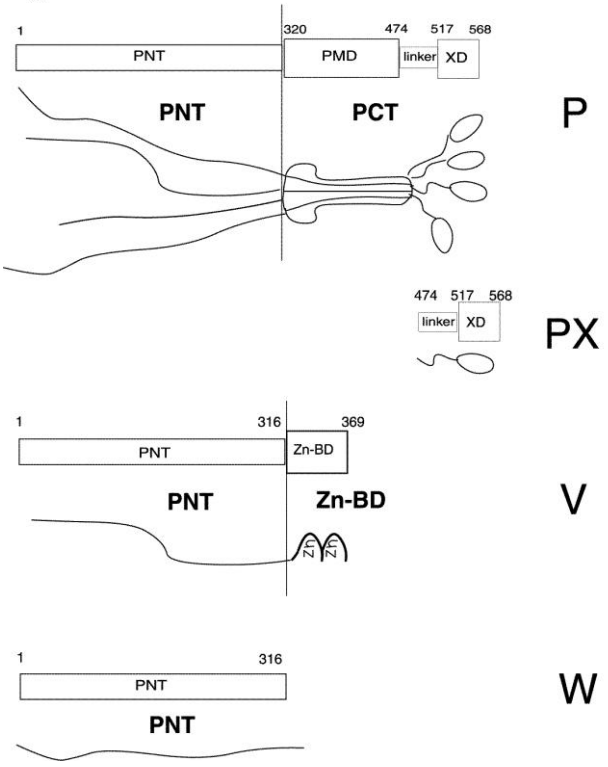

Figure 5. Proteins expressed from the SeV P gene and partially overlapping with P

The SeV P gene encodes 4 proteins that share common domains by a variety of mechanisms (see text). A schematic picture of each protein is shown under its domain representation. Globular regions are represented as large boxes, and unstructured or flexible regions as narrow boxes (see also figure 6). PMD: P multimerization domain. The protein X (PX) consists of a flexible linker and of the X domain (XD). V consists in a zinc-binding domain (Zn-BD) fused downstream of PNT. W is composed of PNT alone.

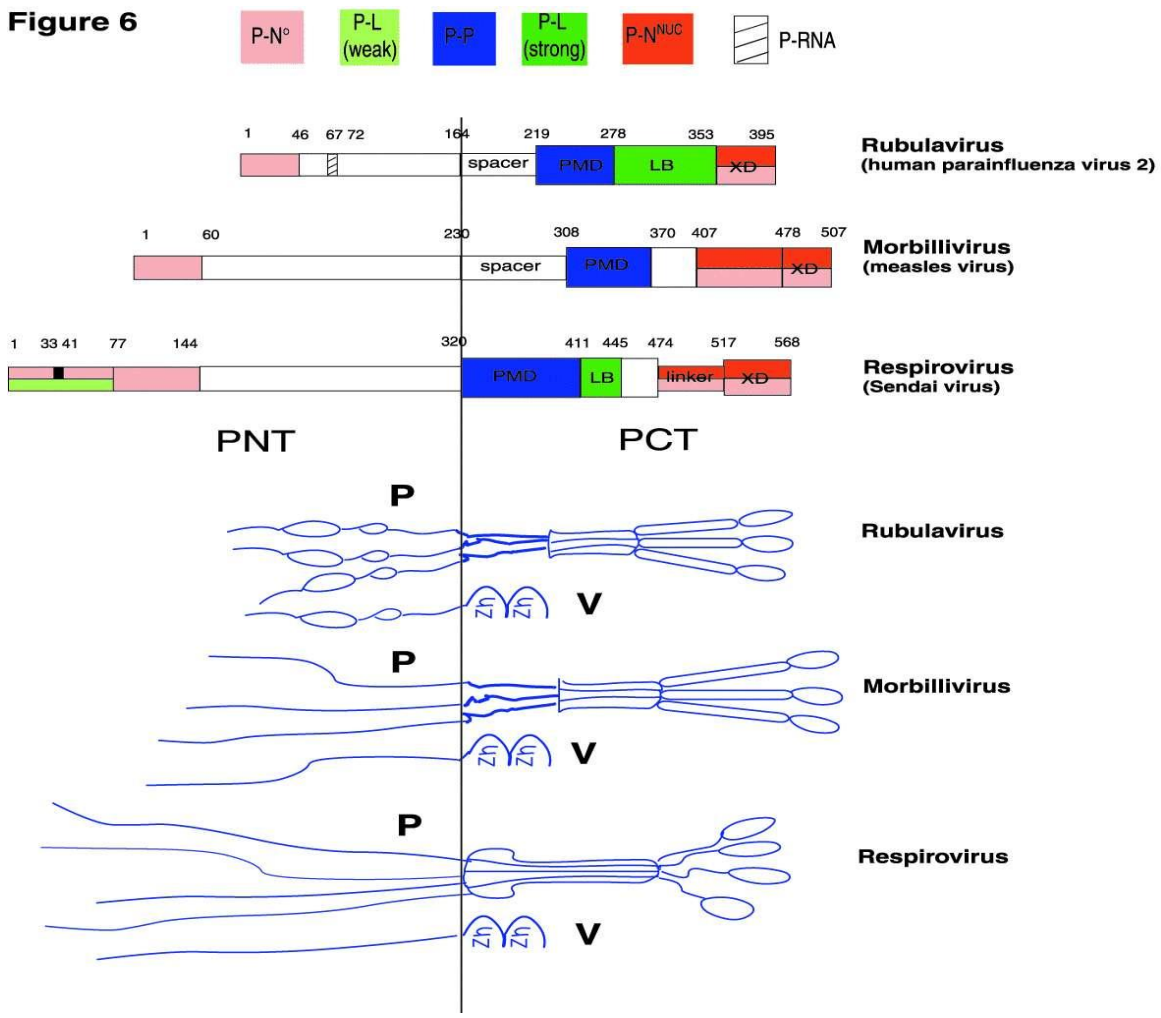

Figure 6. Comparison of the organization of P in *Rubulaviruses, Morbilliviruses,* and, *Respiroviruses*.

The structural and functional organization of P in the 3 genera is presented in the upper panel. A schematic depiction of the corresponding proteins is found in the lower panel.

Upper panel

Globular regions are represented as large boxes, and unstructured or flexible regions as narrow boxes. The regions of interaction between P and itself, or N and L, are identified by a color code (above the diagram). PNT: P N-terminal moiety. PCT: P C-terminal moiety. PMD: P coiled-coil multimerization domain (represented with a dumbbell shape according to (Tarbouriech *et al.*, 2000)). LB: L-binding site. XD: "X domain" (conserved predicted α-helical bundle). The main N°-binding site in SeV P is represented by a black box. The weak L-binding site postulated to take part in a tripartite interaction between N, P, and L (see text) is tentatively represented.

Lower panel

The present study suggests that *Rubulavirus* PNT might contain disordered regions, and it is depicted as a mixture of lines and of contoured forms. The presumably flexible spacer region of *Rubulavirus* and *Morbillivirus* PNT is rendered as a bold wavy line. The region connecting LB and XD is flexible in *Respiroviruses* whereas its status in other genera is unknown. For these genera it has been represented as thin rectangles.

The PCT moiety

Three common features of PCT in *Paramyxovirinae* are: 1) oligomerization through a predicted coiled-coil region; 2) the presence at the extreme C-terminus of a short predicted bundle of three α-helices, which is involved in binding to N (these features were first described in (Curran *et al.*, 1995a)); 3) the presence of a strong L-binding site.

Oligomerization through the coiled-coil region (called PMD, "P Multimerization Domain") has been assessed for phosphoproteins of the three genera (Curran *et al.*, 1995a; Harty and Palese, 1995; Nishio *et al.*, 1997). SeV P is tetrameric; hPIV-2 is trimeric; the status of MV P is unknown. Experimental data available so far suggest that P performs both transcription and replication in its multimeric form. SeV P is a tetramer when complexed with L (Curran *et al.*, 1995b), in its supplementary role (Curran, 1996; Curran, 1998), and in N°-P, with a P/N ratio of 4:1 (Curran, personal communication). It is also a multimer, possibly a trimer, in hPIV3 N°-P (Curran, 1996; Curran, 1998; De *et al.*, 2000). However, a P protein lacking PMD and thus unable to multimerize could still bind L (Curran *et al.*, 1994; Smallwood *et al.*, 1994). In agreement, the studies of De *et al.* suggest that P can bind L as a monomer, although multimerization is necessary for activity (De *et al.*, 2000).

The extreme C-terminal domain of PCT (herein called XD, "X Domain", because it is part of the protein X (see figure 5) is moderately conserved in sequence throughout *Paramyxovirinae* (Curran *et al.*, 1995a). Its deletion abolishes binding to N^{NUC} and to N° (Curran *et al.*, 1995b; Curran *et al.*, 1994; Harty and Palese, 1995; Liston *et al.*, 1995; Nishio *et al.*, 1996; Shaji and Shaila, 1999). In *Respiroviruses* and *Rubulaviruses*, the stable L-binding site (herein called LB) is located between PMD and XD (Curran *et al.*, 1994; Nishio *et al.*, 2000; Smallwood *et al.*, 1999). *Morbillivirus* PCT also contains a stable L-binding site (Liston *et al.*, 1995). Its precise location is unknown but it is also expected to be found between PMD and XD, given the modular nature of P and the sequence conservation in this region among the three genera. L may contact other regions of P than the minimal L-binding site during transcription (Bousse *et al.*, 2001; Bowman *et al.*, 1999).

Detailed structural information is available only for SeV PCT. The structure of a slightly shortened version of PMD has been solved by X-ray crystallography (Tarbouriech *et al.*, 2000). PMD forms a very stable coiled-coil (aa 365-433) buttressed by a bundle of three short α-helices (aa 320-364). The tridimensional structure of the protein X has been solved by NMR (Marion *et al.*, 2001). It is monomeric and consists in a flexible linker (aa 479-517) fused to XD (aa 479-568) (figure 6).

In *Rubulavirus* and *Morbillivirus* PCT (as well as in *Henipavirus* and in NDV PCT) an additional spacer region 55 to 70 aa long is found upstream of PMD (MV:aa 231-300; hPIV2: 165-219) (figure 6). Although there is little general sequence conservation, the composition of this region is peculiar (being rich in G, S, and A) and conserved throughout these viruses

(figure 7). Both spacer regions are depleted in most "order-promoting" residues (see above) and enriched in most "disorder-promoting" residues. Such an analysis is relevant because the regions are long enough (> 40 aa) and modular (Williams *et al.*, 2001). In agreement with these sequence features, PONDR predicts a LDR in the hPIV-2 and MV spacer regions (figure 4C). However, the mean net charge/hydrophobicity ratio of the spacer region is that of a globular domain, although it is borderline (see figure 4B)**.** Thus, while it is likely that this region is at least in part disordered, this demands experimental confirmation. Interestingly, the spacer region overlaps the V ORF (this is what, coupled to its peculiar composition, prompted us to name it "spacer"). Similarly, the C ORF overlaps that of PNT. This suggests that an examination of proteins encoded by overlapping reading frames might reveal unstructured regions, a hypothesis emitted independently by K. Dunker (personal communication).

The SeV P gene encodes a structured region in both overlapping ORFs of PCT and V (helices A to C of PMD (Tarbouriech *et al.*, 2000), and Zn-binding domain, respectively) (figure 6). However, for SeV it is the P reading frame which imposes the codon use of the P gene in the overlapping P/V region, contrary to *Morbilliviruses* and *Rubulaviruses*, in which the V reading frame has priority on the P reading frame (Jordan *et al.*, 2000). Thus, the presence of overlapping reading frames does not necessarily imply the presence of unstructured regions; the occurrence of such regions might be reflected by the codon use within overlapping regions of genes.

A hint that the spacer region might play a functional role in the N-P interaction comes from studies of (Huber *et al.*, 1991). Aa 300-507 of MV P are sufficient for binding the nucleocapsid, but the additional contribution of aa 204-299 (mainly composed of the spacer region) is necessary for proper localization of P and consequently for retaining N in the cytoplasm.

Figure 7

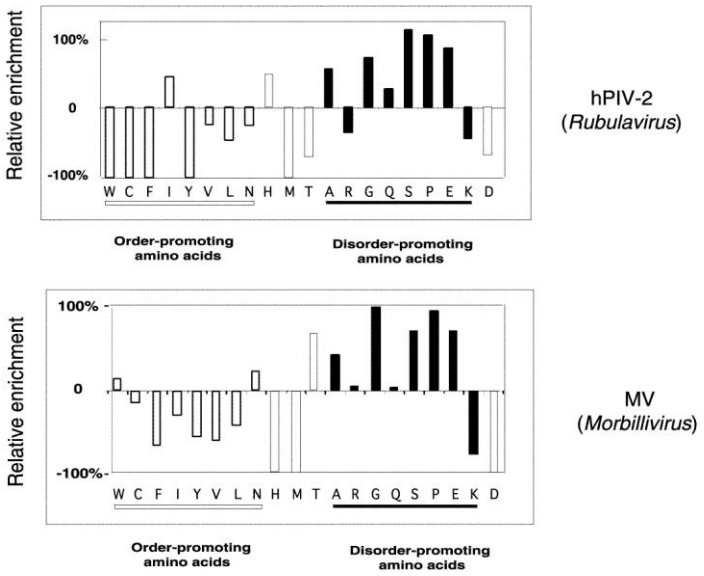

Figure 7. Amino acid composition of the spacer region

Deviation in sequence composition from the average values in the PDB for the spacer regions of hPIV-2 and MV P.

Order-promoting and disorder-promoting amino acids are indicated by empty bars, and by black bars, respectively. Amino acids that are indifferently enriched or depleted in disordered regions of proteins are represented by empty bars with a thin contour. The amino acids are presented in a slightly modified order compared to (Romero *et al.*, 2001), in order to group order-promoting and disorder-promoting residues.

The PNT moiety

As pointed out by Lamb *et al.*, there are functional and structural differences between *Morbillivirus* and *Respirovirus* PNT, which are acidic and 230-320 aa in length, and *Rubulavirus* PNT, which are shorter (about 160 aa), basic, and exhibit a low sequence similarity with the C proteins of *Respiroviruses* (Lamb and Kolakofsky, 2001). Furthermore, *Rubulavirus* V is found in virions, contrary to the V proteins of the other genera (Paterson *et al.*, 1995). However, PNT shares at least one function throughout the three genera, that of binding N° (Curran *et al.*, 1995b; Nishio *et al.*, 1996; Precious *et al.*, 1995; Spehner *et al.*, 1997). *Morbillivirus* and *Respirovirus* PNT contain all phosphorylation sites of P; such an information is not available for *Rubulaviruses* (Byrappa and Gupta, 1999; Byrappa *et al.*, 1996; Das *et al.*, 1995; Hsu and Kingsbury, 1982; Jonscher and Yates, 1997; Vidal *et al.*, 1988). Phosphorylation of *Paramyxoviridae* phosphoproteins is highly conserved, implying functional significance (Lamb and Kolakofsky, 2001). However, its role is still unclear, in particular because different results are obtained depending on the system of study. Phosphorylation of SeV P by protein kinase C-ζ would play a role in activating transcription

(Huntley *et al.*, 1997; Liu *et al.*, 1997) yet the primary phosphorylation site of SeV P can be destroyed without affecting viral replication and pathogenesis (Hu *et al.*, 1999). Further deletion of the secondary phosphorylation sites had no effect on viral replication in cell culture either (Hu and Gupta, 2000). Furthermore, P and N associate in bacteria (where they are not phosphorylated), thus precluding a role for phosphorylation in their assembly (Shaji and Shaila, 1999).

There is no obvious sequence similarity between *Morbillivirus* and *Respirovirus* PNT, as opposed to that noted for PCT. However, both domains share structural and functional features. Recombinant MV PNT is an unstructured monomer, fully extended in solution (Karlin *et al.*, 2002b). Its sequence properties (namely low hydrophobicity and relatively high net charge due to the presence of numerous acidic residues) are typical of intrinsically disordered proteins. PONDR (Li *et al.*, 1999; Romero *et al.*, 1997) predicts more than half of MV PNT to be disordered. The same sequence features are found in all *Morbillivirus* and *Respirovirus* PNT (Karlin *et al.*, 2002b), suggesting that they are a general property of PNT throughout these genera. Experimental support for this thesis comes from older studies on SeV P (Chinchar and Portner, 1981a; Deshpande and Portner, 1985). P bound to nucleocapsids is composed of a 40 kDa C-terminal core resistant to proteolysis, while the remaining N-terminal region (extending to at least aa 221, out of 320 for PNT) is hypersensitive to proteolysis. Thus, SeV PNT is in large part unstructured within the viral RNP. On the other hand, *Rubulavirus* PNT do not have the peculiar hydrophobicity/net charge ratio typical of intrinsically disordered proteins (Karlin *et al.*, 2002b) (figure 4B). This is in agreement with their marked difference with *Respirovirus* and *Morbillivirus* PNT. However, PHD (Rost, 1996) predicts little secondary structure in *Rubulavirus* PNT (4% α-helix and 12% β-sheet), and PONDR predicts LDRs in MuV, SV5, and hPIV2 PNT (figure 4C). Furthermore, there is very little sequence conservation among *Rubulavirus* PNT, another possible indication of structural disorder. Thus, the properties of *Rubulavirus* PNT are markedly different from those of globular proteins. An estimation of the error rate of the hydrophobicity/net charge method can be drawn from (Uversky, 2002). In this work, no globular protein was found to have a ratio located on the left side of the line, indicating that the negative error rate for the prediction of globular proteins must be very low. However, 5 unfolded proteins out of 105 - which were all borderline - were wrongly assigned as being globular, indicating a positive error rate of about 5%. Thus, it is possible that *Rubulavirus* PNT are wrongly assigned as being globular, or that they contain both ordered and disordered regions, although this demands experimental confirmation.

Noticeably, Chinchar *et al.* performed on NDV P studies similar to those carried out on SeV P and also found an organization into an N-terminal region sensitive to proteolysis and a C-terminal core (Chinchar and Portner, 1981b). However, NDV PNT is markedly different from *Rubulavirus* PNT. The aa 1-46, which constitute the N°-binding region of

Rubulavirus PNT (see below), are conserved in sequence among *Rubulaviruses* and NDV. However, the remainder of *Rubulavirus* PNT is very basic, whereas that of NDV PNT is acidic, similarly to *Morbilliviruses* and *Respiroviruses*. Therefore, the structural state of *Rubulavirus* PNT remains undefined.

The functional organization of *Respirovirus* PNT is best known from studies on SeV. It is altogether dispensable for transcription (Chinchar and Portner, 1981a; Curran, 1996). Surprisingly, aa 145-320 (which contain the sites of phosphorylation – see below) are not needed for replication either. A phosphoprotein formed by aa 1-77 fused to PCT can perform replication *in vitro*, indicating that aa 1-77 contain all determinants required for replication, at least *in vitro* (Curran *et al.*, 1994). In particular, the main N°-binding site has been mapped to aa 33-41 (Curran *et al.*, 1995b). aa 78-144 also interact with N and partially prevent its self-assembly (Curran, 1996; Curran *et al.*, 1995b; Curran *et al.*, 1994; Horikami *et al.*, 1996). Likewise, the extreme N-terminus of *Morbilliviruses* and *Rubulaviruses* PNT is necessary for binding to N° (Harty and Palese, 1995; Nishio *et al.*, 1997; Nishio *et al.*, 1996; Shaji and Shaila, 1999; Watanabe *et al.*, 1996).

Studies on V have helped to clarify the role of PNT. The protein V of the *Morbillivirus* RPV interacts with L, via PNT, as proved using yeast two-hybrid and immunoprecipitation experiments (Sweetman *et al.*, 2001). No such direct evidence has been provided for SeV, but several arguments indicate that *Respirovirus* L and PNT also interact. SeV V and W inhibit replication, an effect that can be relieved by overexpression of L (Curran *et al.*, 1991). Furthermore, a chimerical bPIV-3 PNT / SeV PCT protein can function in SeV replication, but is very sensitive to the amount of L (Curran *et al.*, 1995b).

It is not known whether PNT has the same conformation within V and P. An antibody raised against aa 65-87 of P recognized MV P but not V, indicating either that PNT has different conformations within these proteins, or that a cellular protein bound to V interferes with the recognition of this region [[Wardrop, 1991 #38]87]. Liston *et al.* reported that V interacts with three cellular partners, PNT and the zinc-binding domain of V being both required for this interaction (Liston *et al.*, 1995). In contrast, the zinc-binding domain alone appears to mediate the interaction of V with DDB-1 (Lin *et al.*, 1998), or the effect of V on pathogenicity (Huang *et al.*, 2000).

In contrast to other genera, *Rubulavirus* V and P bind RNA through a stretch of basic residues around aa 70. The functional significance of this observation is unknown.

An exception to the mapping of P/N/L interactions reported above was found for porcine *Rubulavirus* (LPMV) (Svenda *et al.*, 2002). LPMV PNT binds L, and it is the zinc-binding region of V (rather than PNT) which binds N. Given the very high sequence conservation of the zinc-binding region, these results are unexpected.

THE LARGE POLYMERASE SUBUNIT (L)

L is thought to be a multifunctional enzyme, carrying most or all catalytic functions necessary for synthesis of viral RNA. Such a large enzyme (about 250 kDa) is almost certainly organized into different domains. However, we know little about the functional organization of L and nothing about its structural organization. This situation stems mostly from two facts: the rarity of L in virions, and its large size that renders recombinant expression difficult. Biocomputing studies have shown that *Mononegavirales* L contain 6 conserved regions (called I to VI), which were proposed to specify the essential activities common to all L proteins (Poch *et al.*, 1990; Sidhu *et al.*, 1993). It is possible that multiple domains contribute to the different steps in viral RNA synthesis, since mutants in different domains gave the same defective phenotypes (Chandrika *et al.*, 1995; Cortese *et al.*, 2000; Feller *et al.*, 2000; Horikami and Moyer, 1995; Smallwood *et al.*, 1999). Ferron *et al.* have recently shown that a domain encompassing the region VI, which was postulated to contain an ATP-binding site, in fact specifies a 2'O methyltransferase domain (aa 1755-1960 of MV L), probably involved in the capping of viral messengers (Ferron *et al.*, 2002).

L must form of a complex with P (through PMD) in order to be stable and active (Horikami *et al.*, 1994). In *Paramyxovirinae,* the stable P-binding site is located in the N-terminal moiety of L (Holmes and Moyer, 2002; Horikami *et al.*, 1994; Parks, 1994). A different region of L also forms a complex with C, which inhibits RNA synthesis (Cadd *et al.*, 1996; Curran *et al.*, 1991; Grogan and Moyer, 2001; Horikami *et al.*, 1997; Reutter *et al.*, 2001). Direct structural studies are necessary to further enlighten the organization of L.

PROTEIN INTERACTIONS WITHIN THE REPLICATIVE COMPLEX

What distinguishes negative-stranded RNA viruses from other viruses is that the nucleoprotein is not a simple structural component, serving merely to package viral RNA. Rather, it plays an active role in RNA synthesis. The components of the viral replication machinery, namely P, N and L, engage in an elaborate macromolecular ballet that is very poorly understood at present, in part due to the large size of the participants. However, the roles of the N°-P and N^{NUC}-P complexes are beginning to appear.

THE N-P COMPLEXES

As mentioned above, N and P form two different complexes: a soluble, cytosolic complex (N°-P), which is the substrate used by the polymerase to encapsidate neosynthesized RNA, and a complex in which P is bound to nucleocapsids (N^{NUC}-P). Elucidating the roles of these complexes is central to our understanding of the viral cycle, and recent advances in mapping the interacting regions of N and P deserve a separate chapter. A common theme emerging from studies on *Paramyxovirinae* is that while an interaction between C-terminal regions of N and P is involved in both N^{NUC}-P and N°-P, the

N°-P complex involves an additional interaction between N and the N-terminus of P (figure 8).

N^{NUC}-P

SeV viral RNPs are composed of about 2600 N, 300 P, and 30 L molecules (Lamb and Choppin, 1977). P is distributed randomly on nucleocapsids isolated from virions whereas it is regrouped in clusters on transcriptionally active nucleocapsids (Portner and Murti, 1986; Portner *et al.*, 1988). The localized distribution of P is not due to a limited number of nucleocapsid tails available for binding N, since it is possible to add exogenous P to nucleocapsids (Ryan *et al.*, 1990). Very little is known about the role of P in RNA synthesis, besides serving as a bridge between L and the (N:RNA) template. For instance, it is not known whether P binds nucleoproteins within the same turn of the helical nucleocapsid or within different turns (see figure 9). About 120 molecules of P (instead of 300) would be expected if tetrameric P were found only in complex with L. Indeed, while L is always associated with P, uncomplexed P molecules are found on nucleocapsids (Portner *et al.*, 1988). In agreement, it was recently demonstrated that supplementary molecules of P, which are not associated with L, are needed in stoechiometric amounts to sustain the activity of the SeV P-L polymerase complex. The N-binding region of P, PX is required for this supplementary role, as well as the stable L-binding region. PNT, on the other hand, is totally dispensable for the supplementary role, and for transcription in general (Curran, 1996). What maintains (P-L) and supplementary P molecules grouped together on the nucleocapsid is also unknown. Apparently, supplementary, free P molecules are not exchanged with the P molecules associated with L (Bowman *et al.*, 1999; Curran, 1996; Curran, 1998; Ryan *et al.*, 1990; Tuckis *et al.*, 2002).

Elegant studies have shown that it is possible to separate the role(s) of P within the P-L complex from its supplemental role, using site-directed mutagenesis (Bowman *et al.*, 1999; Tuckis *et al.*, 2002). Some of the variant phosphoproteins that were inactive in transcription could be rescued by wild-type P, indicating that they were defective only in the supplemental role and not in their role within P-L. The reciprocal phenotype was also obtained (i.e. a variant P that could perform a supplemental role but was inactive within the (P-L) complex, and could not be rescued by wild-type supplemental P).

The precise nature of this supplementary function of P is unknown. Supplementary P (or indeed, P bound to L) might help expose the RNA by locally opening the nucleocapsid coil (Curran, 1998). This hypothesis originates in part from the findings that viral nucleocapsids are found not only in a tight helical conformation but also in looser helical forms; these changes are reversible depending on salt concentrations (Egelman *et al.*, 1989; Heggeness *et al.*, 1981). However, recent studies on VSV indicate that the base moieties (but not the phosphate moieties) of encapsidated RNA are still accessible to the solvent, and thus it is

conceivable that the nucleocapsid never actually uncoils during RNA synthesis (Iseni *et al.*, 2000).

Two discontinuous regions of PCT, PMD and PX, are necessary and sufficient for binding to N^{NUC} (Ryan *et al.*, 1991; Ryan and Portner, 1990). The exact role of PMD in binding N^{NUC} is still a matter of discussion. It could be associated with PX in binding N^{NUC} or it could play an indirect role, providing multimerization required to stabilize weak PX-N_{TAIL} interactions, as proposed by Curran *et al.* (Curran, 1998) (see figure 8). Mixed P tetramers containing only 3 PX domains are able to bind N^{NUC} but are not active in transcription *in vitro*. This opens a possible role for the free, unbound PX region(s), such as cartwheeling on the nucleocapsid, continuously opening and breaking N-PX liaisons (Curran, 1998; Curran and Kolakofsky, 1999). In contrast De *et al.* recently showed that phosphoproteins of hPIV-3, a close relative of SeV, containing less than 4 PX domains could perform transcription *in vivo* (De *et al.*, 2000). Thus, the role of multimerization of P in binding to N^{NUC} remains undefined.

Figure 8

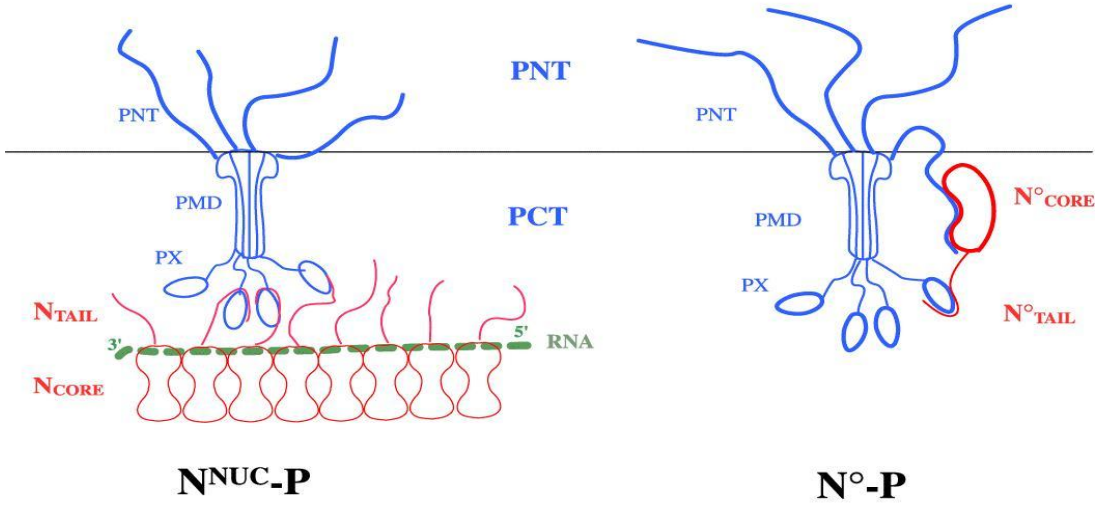

Figure 8. The N°-P and N^{NUC}-P complexes

We have represented the N°-P and N^{NUC}-P complexes of SeV, a *Respirovirus*, using the same conventions as in figures 3 and 6.

N^{NUC}-P

P is shown bound to N^{NUC} through 3 of its 4 C-terminal PX "arms", as in the model of (Curran and Kolakofsky, 1999) (see text). N^{NUC} contains RNA, shown as a green dotted line near the P-binding site, by analogy with RV N, in which the RNA is thought to reside near the P-binding site (Schoehn *et al.*, 2001).

N°-P

SeV N°-P has a 1:4 stoechiometry (J. Curran, personal communication). N° is tentatively depicted as being bound through its tail to one PX "arm", while the core of N° is bound to one PNT terminal region. The drawing emphasizes the potential importance of the flexibility of PNT and N_{TAIL}. It is not known, however, if these latter regions fold upon binding their respective partners.

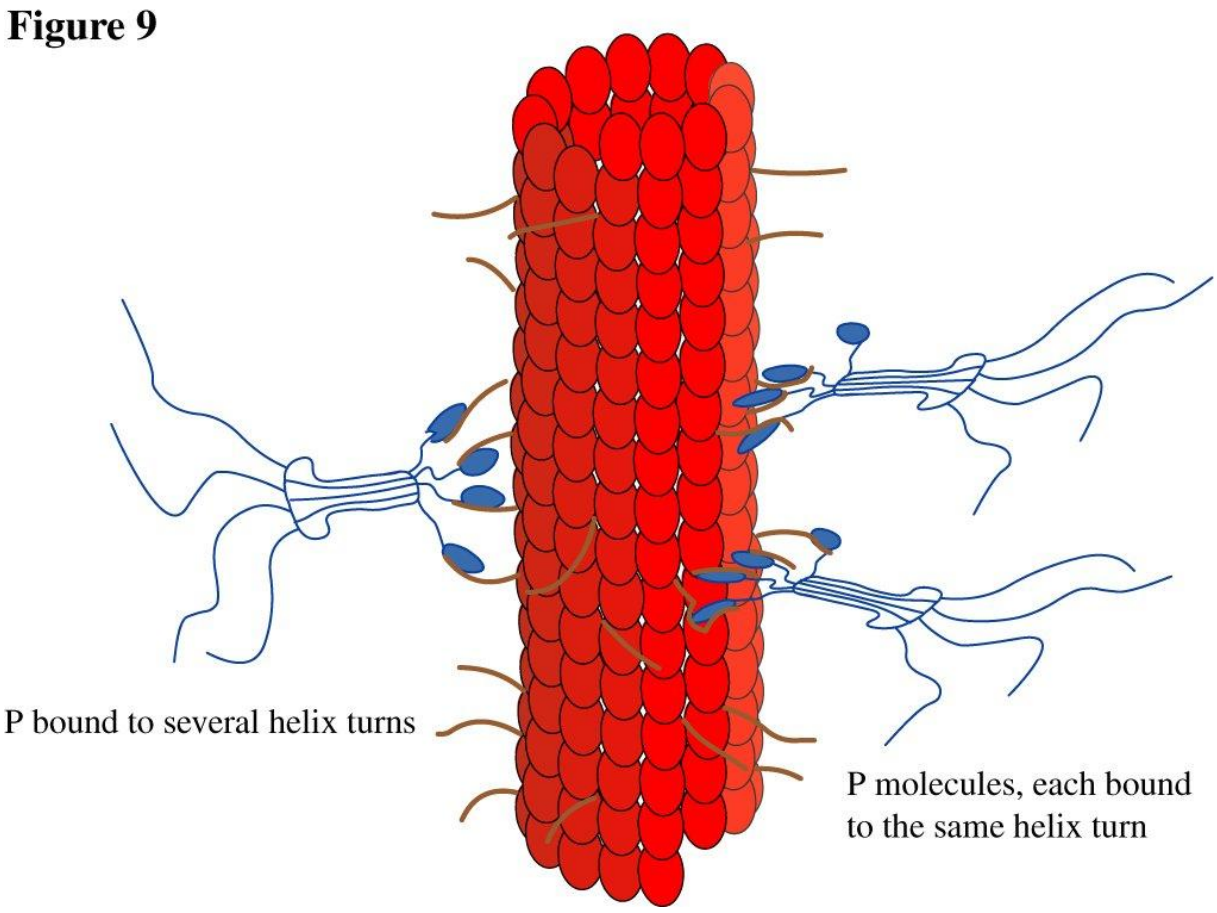

Figure 9. Binding of P to the nucleocapsid

N is represented as a helical nucleocapsid with 13 N subunits per turn (red ovals), like that of SeV. The red ovals represent N_{CORE}, while the brown lines represent N_{TAIL}. Only a few N_{TAIL} regions are represented for clarity. The drawing is conceived to emphasize the structural flexibility of P, as well as our ignorance of the precise modalities of N-P binding. During viral RNA synthesis, P could be bound either to the same turn of the helix (right part of the figure) or to different turns (left part). The flexibility of P probably gives it a considerable reach.

N°-P

For *Morbilliviruses* and *Respiroviruses*, as well as for NDV, co-expression of N and P leads to the formation of a soluble, cytosolic N°-P complex (Curran *et al.*, 1995b; Errington and Emmerson, 1997; Spehner *et al.*, 1997), whereas V only partially prevents self-assembly of P (Curran *et al.*, 1995b). In contrast, co-expression of *Rubulavirus* N and P leads to the formation of large intracellular cytoplasmic aggregates, presumably composed of nucleocapsids bound to P, while co-expression of N and V prevents formation of such aggregates (Nishio *et al.*, 1996; Precious *et al.*, 1995; Randall and Bermingham, 1996; Watanabe *et al.*, 1996). Thus, in *Rubulaviruses* it is probably V that keeps N in a soluble form.

Studies on SeV suggested that an N°-P complex is absolutely necessary for the polymerase to initiate encapsidation but that in the presence of large amounts of soluble N alone, the polymerase can carry out replication of preinitiated chains (Baker and Moyer, 1988). Therefore, formation of a complex of N° with either P or V would have at least two separate functions:

i) prevent illegitimate self-assembly of N. This role is fulfilled by V in *Rubulaviruses* and by P in other genera.

ii) allow the polymerase to deliver N to the nascent RNA to initiate replication, a role devoted to P in *Morbilliviruses* and *Respiroviruses,* and perhaps also in *Rubulaviruses* (Randall and Bermingham, 1996).

Deletion of either N_{TAIL} or PX prevents the formation of a stable N°-P complex as well as that of an N^{NUC}-P complex (Curran *et al.*, 1993; Curran *et al.*, 1995b), which strongly suggests that N_{TAIL} and PX interact in both complexes. However, the residues of PX involved in binding N° and N^{NUC} are not identical and can be separated using site-directed mutagenesis (Tuckis *et al.*, 2002). At first glance, the location of N-binding sites in P seems different between *Morbilliviruses* and *Respiroviruses*. Studies on *Morbillivirus* P-N interaction were performed using the highly sensitive yeast two-hybrid system, whereas those on SeV were carried out using immunoprecipitation. In contrast with SeV P, the extreme C-terminus of *Morbillivirus* P could be deleted without abolishing the N-P interaction, since RPV P containing only the N-terminal 59 aa and PMD could interact with N. In further contrast with *Respiroviruses*, no interaction between N and PCT could be detected (Harty and Palese, 1995; Shaji and Shaila, 1999). However, MV N and PCT clearly interact (Huber *et al.*, 1991). Therefore, it would seem that an artifact in the yeast two-hybrid system prevents detection of one type of N-P interaction (presumably the N^{NUC}-P interaction). Consistent with this interpretation is the fact that the yeast two-hybrid system does not allow detection of the N-N interaction (Shaji and Shaila, 1999), probably because fusion of N with large proteins prevents its proper self-assembly (Warnes *et al.*, 1995). Using the yeast two-hybrid system to

study N-P interactions of hPIV-3, a close relative of SeV, Zhao *et al.* found results very similar to those obtained in *Morbilliviruses* but diverging from immunoprecipitation studies on SeV (Zhao and Banerjee, 1995). Thus, although the yeast two-hybrid system provided some interesting results, due to its great sensitivity, more studies distinguishing between N° and N^{NUC} are needed to clarify the respective roles of PX and PMD.

Deletion studies are impractical to map the regions of N_{CORE} involved in N°-P complex formation, because N_{CORE} does not appear to have a modular structure, and consequently it is difficult to distinguish between gross structural defects and specific effects of deletions. Deletions in most regions of N_{CORE}, with the exception of aa 189-239, abolish all N-P interaction (Bankamp *et al.*, 1996; Liston *et al.*, 1997). The only direct study is that of Gombart *et al.* (Gombart *et al.*, 1993). The authors showed that for MV, there are two soluble forms of N present in the free pool of proteins: the neosynthesized nucleoprotein (herein called (n)), and N°. By analogy with SeV, it is assumed that (n) is present in the free pool of proteins as a monomer (Baker and Moyer, 1988). (n) undergoes several changes in conformation before being incorporated in nucleocapsids. In certain infected cells, or in pathological cases of measles, (n) can go to the nucleus where it forms inclusions (reviewed in (Griffin, 2001)). This migration to the nucleus is prevented by P through formation of a cytosolic (N°-P) complex. N° is then rapidly incorporated in nucleocapsids (Gombart *et al.*, 1993; Huber *et al.*, 1991; Spehner *et al.*, 1997) (figure 10). A change of conformation of (n) creates an epitope recognized by a conformational antibody (Gombart *et al.*, 1993). This epitope (formed by aa 23-239) is not created by the assembly of N into nucleocapsids, since it is present in both N° and N^{NUC}. Association of P with N° (but not with N^{NUC}) interferes with the recognition of this epitope, indicating that aa 23-239 of N are involved in N°-P but not N^{NUC}-P binding. A similar maturation process of the Pneumovirinae hRSV N has been described (Murray *et al.*, 2001). At present it is not known whether it is association of P that drives the change in conformation of (n) into N°, or if N° pre-exists before association with P, in infected cells. However, the presence of P is not necessary for formation of this epitope *in vitro*, since the denatured polypeptide 23-239 can refold to yield the epitope.

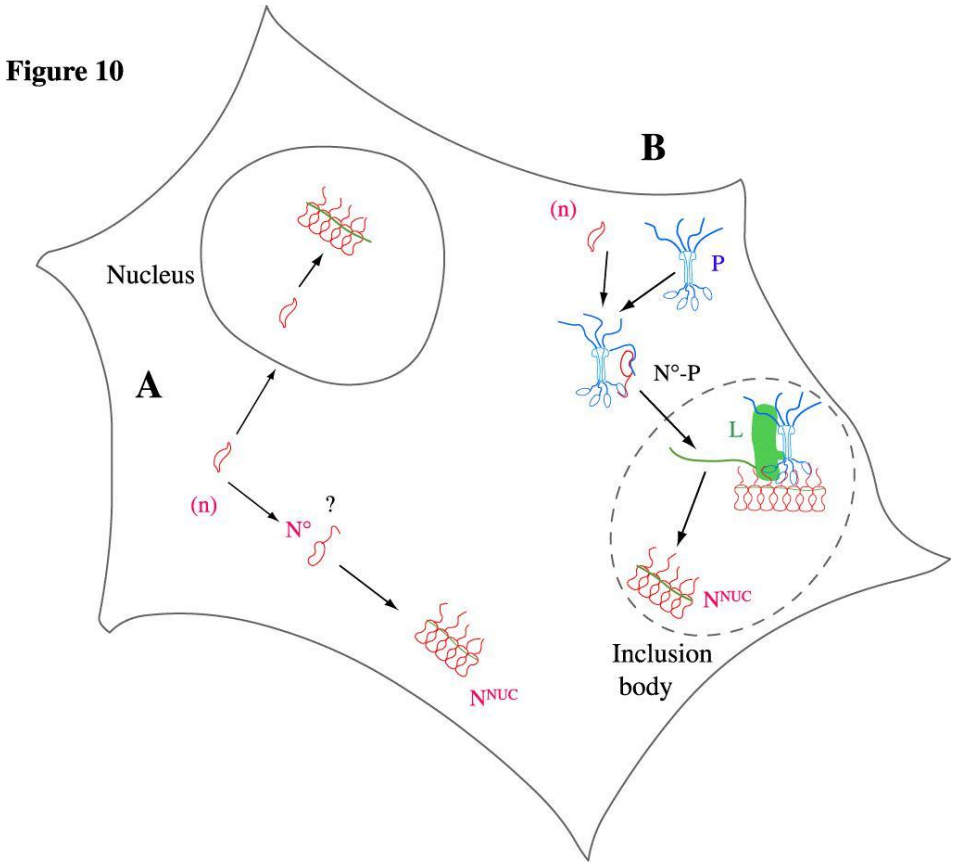

Figure 10. Maturation of MV N

Interior of an infected cell, with the nucleus (top left), and a cytoplasmic inclusion body (bottom right). The nucleoprotein (red) is found under several conformations. The neosynthesized form, called (n), and a more mature form, N°, are both cytosolic, whereas the assembled form, N^{NUC}, is probably bound to the cytoskeleton (which is not represented). The polymerase complex is formed of L (light green) and one tetramer of P (blue) (see also figures 8 and 11). This representation of the maturation of N is derived from the studies of (Gombart *et al.*, 1993).

A. In the absence of P, (n) can self-assemble illegitimately on cellular RNA (bottom). It can also migrate to the nucleus, where it forms nucleocapsid-like particles. It is not known whether it undergoes a change in conformation directly from (n) to N^{NUC}, or whether N° is an intermediate conformation in the process.

B. P forms a complex with N°, which prevents its illegitimate encapsidation. This N°-P complex is used by the polymerase to encapsidate neosynthesized RNA during replication (which takes place in cytoplasmic inclusion bodies)

A TRIPARTITE N°-P- (L-P) COMPLEX?

A transient, tripartite complex among the unassembled N°, P, and the polymerase (P-L) devoted to encapsidate nascent RNA during replication had been proposed in (Curran *et al.*, 1995b). In our opinion, recent evidence provides strong support for this hypothesis. The most direct support comes from mutation studies of SeV P, centered around PX (Tuckis *et al.*, 2002). Two variants of P, JT2 and JT7, can still bind N^{NUC} and L, and are active in transcription. However, they no longer form a stable N°-P complex. Amazingly, these mutants are capable of supporting *in vitro* replication of a defective-interfering virus only when N, variant P, and L are all co-expressed, but not when cellular extracts with co-expressed (N, variant P) and (variant P,L) are provided separately. The latter assay is called "*in vitro* reconstitution of replication".

These data can be interpreted by proposing that the mutations have separated 2 functions of P in replication: i) that of preventing N from self-assembling illegitimately, and ii) that of delivering N° to the polymerase. Failure of the *in vitro* reconstitution of replication assay can then be easily explained if P has lost the preventive (chaperone) function but not the delivery function.

Indeed, the variant P cannot form a stable N°-P complex, and consequently most N is found as N^{NUC} in the N+P extract, unable to serve as an encapsidation substrate for the polymerase complex (Tuckis *et al.*, 2002). In contrast, when N,P, and L are co-expressed, PNT, which has both an N° and a (weak) L binding site, could bring together L and N°, before the latter self-assembles. The polymerase would subsequently use N° for encapsidation (figure 11). In this model, the variant phosphoproteins have kept their function of putting in contact N° and the polymerase, located within PNT, which has not been mutated. Lastly, when none of these functions of P are present (for instance when P is not provided in the N extract), no replication occurs (Horikami *et al.*, 1992).

This model can also account in a very similar way for the fact that SeV N_{CORE} can be used by the polymerase for encapsidation, but only if it is co-expressed with P and L (Curran *et al.*, 1993). N_{CORE} does not form a stable N°-P complex, since it lacks the N_{TAIL} region. However, it can still interact, although weakly, with PNT (Curran *et al.*, 1993), which would allow its delivery. Thus, mutations of either the PX domain or of its cognate N_{TAIL} can separate the preventive function of P° from its delivery function.

Clearly, the function of P in allowing the polymerase to deliver N to the nascent RNA is not restricted to an interaction with L through PNT. Rather, P must contain a third, additional function in replication, which maps at least in part to PCT. Otherwise, V (which interacts with N and L, see above) should be able to substitute for P in the N°-P complex, whereas on the contrary it has an inhibitory effect on replication (Curran *et al.*, 1991; Curran *et al.*, 1992).

This third function does not reside in the L-binding domain of P, because P lacking this region can still form an N°-P complex functional in replication (Curran *et al.*, 1994). This function might be that of binding the nucleocapsid template. In this hypothesis, tetrameric P, which is bound to one N° molecule, could still bind the template N^{NUC} through its 3 free PX regions, which might put N°-P in the proper position to interact with the polymerase (see figure 11). In agreement, the P variants JT2 and JT7, mentioned above, can still bind N^{NUC} (Tuckis *et al.*, 2002).

The model of tripartite P-N-polymerase interactions would accommodate previous suggestions concerning the role of V in interfering with replication: it might be either by preventing N°-P complex formation, or by interfering with the use of N°-P (Curran *et al.*, 1995b; Horikami *et al.*, 1996). V probably acts at both levels. On the one hand, overexpression of SeV L can overcome the inhibitory effect of Sev V/W on replication (Curran *et al.*, 1991), which suggests interference at the L-PNT level. On the other hand, MV V can compete with the N°-P interaction (Tober *et al.*, 1998). Thus, V probably also interferes with the formation of the N°-P complex.

Finally, the mapping of N-P-L interactions in *Rubulaviruses* provides further support to the hypothesis of a tripartite N°-P-(L-P) complex. Ostensibly, the unstructured tail of hPIV-2 N contains both a P/V binding site for N° and an L-binding site, which overlap (see figure 3). Clearly, when L contacts N° it could also contact P (more precisely, PNT) very easily.

In summary, P has at least 3 functions within N°-P:
i) **preventing illegitimate self-assembly of N**. This function is mainly localized within aa 1-144 of SeV PNT, with a major contribution of aa 33-41, but requires the participation of PCT.
ii) **delivering N to the polymerase**, which can be further subdivided into
iia) **interaction between N°, PNT and L**. This function probably resides within residues 1-77 of PNT, since aa 78-320 of SeV P are altogether dispensable for replication.
iib) a third, unknown function which requires the nucleocapsid-binding and the multimerization domains but not the polymerase-binding domain. This function might involve **binding to the (N:RNA) template**.

An animated model of replication featuring the postulated tripartite complex is presented on our Web site (http://afmb.cnrs-mrs.fr/…/≈dkarlin/model.html).

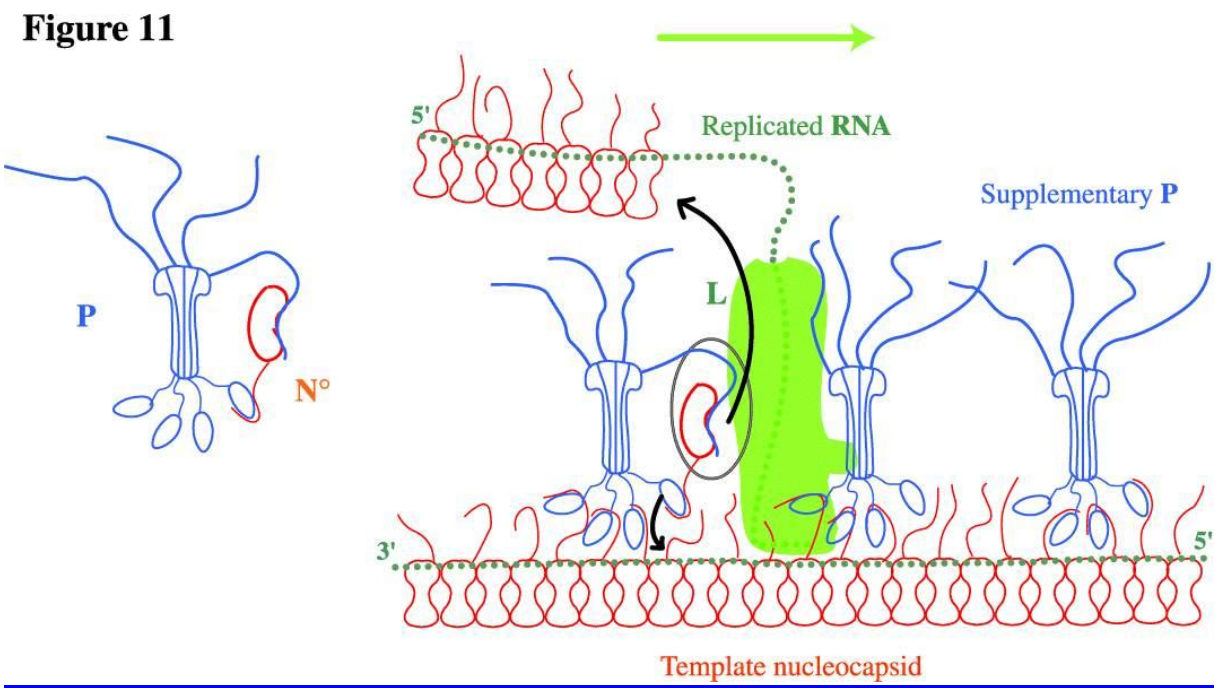

Figure 11. Tripartite N°-P-L interaction during replication

Respirovirus polymerase complex actively replicating genomic RNA. L (large green form) is bound to one P tetramer at the C-terminal end of PMD. A supplementary molecule of P that is not bound to L is represented to the right. Neosynthesized RNA can be seen emerging from the polymerase. It is already partially encapsidated, although there is no evidence that encapsidation begins at the extreme 5' end, as we have represented it. An N°-P delivery complex is put into contact with the newly replicated (N:RNA) by L through a PNT-L contact, forming a tripartite complex N°-P-polymerase (circled). P within N°-P is represented as being bound to the *template nucleocapsid* through its free PX "arms", although there is no experimental evidence to support this hypothesis. P complexed to N° probably does not bind *newly replicated (N:RNA)*, however, since tailless N can function in encapsidation (Curran *et al.*, 1993). It is tempting to speculate that the proximity of the template nucleocapsid might cause the tail of N° to be released by PX. This would result in the incorporation of N° into the nascent N:RNA chain (long arrow), as well as in binding of the free PX to the N:RNA template (short arrow). See also figure 8 for a comparison of the N°-P and N^{NUC}-P complexes.

STRUCTURAL FLEXIBILITY WITHIN THE REPLICATIVE COMPLEX

In the present review, we have shown that the phosphoproteins and nucleoproteins of *Paramyxovirinae* probably contain other unstructured regions than those already characterized. Likewise, the N-terminal domain of *Rhabdoviridae* P proteins such as those of RV or VSV have the net charge/hydrophobicity ratio typical of intrinsically disordered regions, and are predicted to be in large part disordered by PONDR (Karlin *et al.*, 2002b). Furthermore, EM studies have shown that RV P is bound flexibly to N_{CORE} through N_{TAIL} (Schoehn *et al.*, 2001). These observations suggest that structural flexibility might be a general property of the replicative complex of non-segmented, negative-stranded viruses.

In these viruses, the only unstructured region that has been characterized by structural methods is MV PNT. It is fully extended in solution *in vitro*, with a hydrodynamic radius (Stokes radius, R_S) of about 4 nm (Karlin *et al.*, 2002b). However, R_S reflects a mean dimension; the full extension of PNT is probably considerably higher (greater than 40 nm). In comparison, one turn of the nucleocapsid is 18 nm in diameter and 6 nm high. Thus PNT could easily stretch over several turns of the nucleocapsid, and since P is multimeric, N°-P might have a considerable extension (figure 9 and 11). In the same vein, it is striking that the N-binding domain of SeV P, which interacts with the intrinsically disordered tail of N, should comprise a flexible linker. This certainly suggests the need for a great structural flexibility. Perhaps this flexibility is necessary for the tetrameric P to bind several turns of the helical nucleocapsid? Indeed, the promoter signals for the polymerase are located on the first and the second turn of the SeV and SV5 nucleocapsids (Murphy *et al.*, 1998; Murphy and Parks, 1999; Tapparel *et al.*, 1998).

One advantage of unstructured regions of proteins might be their ability to bind to a wide variety of structurally-distinct substrates (Dunker *et al.*, 2001; Wright and Dyson, 1999). The pattern of interactions of N_{TAIL}, and that of PNT (noted by Curran *et al.* as early as 1994 (Curran *et al.*, 1994)) are consistent with this hypothesis. For instance, after lysis induced by MV, cells release viral nucleocapsids, which bind the human immunoglobulin G receptor FcγRII through N_{TAIL}. Binding of N_{TAIL} alone to FcγRII is sufficient to induce partial immunosuppression (Marie *et al.*, 2001; Ravanel *et al.*, 1997). Thus, MV N_{TAIL} takes part in numerous interactions (with P (within both N°-P and N^{NUC}-P), FcγRII, and probably M). Likewise, PNT interacts not only with N° and L, but also with several cellular proteins (Liston *et al.*, 1995). There is a striking parallel between the N_{TAIL}-PX interaction and the PNT-N_{CORE} interaction. Both interactions are not stable by themselves and must be strengthened by the combination of other interactions; these properties are reminiscent of the acidic activation domains (AADs) of cellular transcription factors (Melcher, 2000). Besides a great flexibility,

this might ensure easy breaking and reforming of interactions, as in the model in (Curran, 1998).

There are intrinsically disordered proteins that carry out their function while being disordered from end to end (Dunker *et al.*, 2001). However, a majority fold in the presence of their ligand(s) (Dyson and Wright, 2002; Uversky, 2002), a process called "induced folding". It will be technically challenging to see whether unstructured regions of N and P undergo such an induced folding. In that respect, we note that the extreme N-termini of *Morbillivirus* phosphoproteins, which are involved in binding N° (Harty and Palese, 1995; Shaji and Shaila, 1999), contain several predicted α-helices. One α-helix could be induced in the presence of the solvent TFE (trifluoroethanol), which is postulated to mimic hydrophobic protein-protein interactions (Karlin *et al.*, 2002b).

Most disordered regions presented here, predicted or characterized, have at least one known partner or function. An exception is the region encompassing aa 400-430 of SeV N, since nucleocapsids lacking this region lose template function, although they still bind P. Thus, it is conceivable that these residues have a role such as a flexible linker/spacer, or binding an additional factor (Buchholz *et al.*, 1994).

The shortest nucleoprotein tail is found in *Pneumovirinae* such as bRSV (Khattar *et al.*, 2000) or hRSV N (Murray *et al.*, 2001). The nucleoproteins of these viruses (391 aa) are organized into a core sufficient for encapsidation (aa 1-364), and a 28-aa C-terminal tail that is not required for encapsidation. In both viruses, a 30-aa region located immediately upstream of the tail is required for interaction with N. This organization is reminiscent of *Rubulavirus* N in which the N^{NUC}-P binding region is located in the 30 C-terminal aa of N_{CORE}. Thus, in both viruses the P-binding site of N is in a region that is exposed but presumably not flexible. Therefore, to the best of our knowledge, the flexible tail present in all *Paramyxoviridae* N is apparently not always required for N^{NUC}-P interaction. One may then ask, what are the common point(s) in the pattern of replication of *Paramyxovirinae* that would require structural flexibility? We believe that there might be at least two. *Paramyxoviridae* M is believed to be the central organizer of virus assembly (Coronel *et al.*, 2001). This role involves polymerization of M as well as binding to N_{TAIL} and probably to other unknown factors in the viral envelope. Strikingly, all the known structures of M proteins reveal substantial flexibility (Gaudier *et al.*, 2001), and the M-N interaction during virus assembly might also require flexibility of N_{TAIL}. The second requirement for flexibility might concern the hypothesized transient N°-P-(L-P) complex that would deliver N to the nascent RNA during replication. In *Respiroviruses* and *Morbilliviruses*, PNT (which has an extended conformation (Karlin *et al.*, 2002b)) would put into contact L and N°. In *Rubulaviruses*, N_{TAIL} (which is not globular) would put into contact L and PNT. In both cases it might be the extended conformation of the domain involved that allows the coming together of the large protein L (250 kDa) with its partners.

A role for phosphorylation?

Zetina has recently shown that a number of intrinsically disordered proteins share a common motif, (LS/SL)(D/E)(D/E)(D/E)X(E/D), called "helix-unfolding motif", always located at the N-terminus of an α-helix (Zetina, 2001). The author speculated that this signal sequence might control the unfolding of these proteins in response to cellular events. Indeed, he noted that this motif (which was degenerated in certain proteins) was similar to the consensus motif of casein kinase II and proposed that phosphorylation might play a role in regulating the structural state of intrinsically disordered proteins. Casein kinase II phosphorylates 3 serine residues of MV P *in vitro* (S86, S151, and S180) (Das *et al.*, 1995). Only the latter, S180, is located at the N-terminus of an α-helix (181-197) predicted using PHD (Rost, 1996). The sequence SLAIEE, which more closely matches the consensus motif defined by Zetina is found upstream of the predicted α-helix encompassing aa 27-38 of MV P; noticeably, we found that this region became structured in the presence of trifluoroethanol (see above) (Karlin *et al.*, 2002b). However, one must be careful, because the tantalizing function(s) of phosphorylation of P has escaped all studies so far. Nevertheless, although speculative, Byrappa *et al.* provided a hint that phosphorylation might play a role in regulating the conformation of P (Byrappa *et al.*, 1996). The authors showed that all potential phosphorylation sites of Sev P were equally accessible to kinases, whereas once phosphorylated, they had a different accessibility to phosphatases.

Remarkably, phosphorylation of SeV N occurs within N_{TAIL}, and that of P occurs within PNT (Byrappa and Gupta, 1999; Byrappa *et al.*, 1996; Das *et al.*, 1995; Hsu and Kingsbury, 1982; Jonscher and Yates, 1997; Vidal *et al.*, 1988). Further studies will tell whether the occurrence of phosphorylation on disordered regions of viral proteins is coincidental, or whether it is a widespread property, as suggested in (Dunker *et al.*, 2001). In that respect, we know of at least one other modular viral phosphoprotein, NSP3 of Semliki forest virus (an *Alphavirus*), which has a C-terminal region that is phosphorylated, has the sequence characteristics of intrinsically disordered proteins and is hypervariable in sequence.

It is tempting to compare phosphorylation of N_{TAIL} with hyperacetylation of histones, which renders the nucleosome less rigid (Dunker *et al.*, 2001). Both post-translational modifications increase the negative charge of the protein. A possible influence of N_{TAIL} on the conformation of N is indicated by the fact that proteolysis of N_{TAIL} renders nucleocapsids more rigid (Heggeness *et al.*, 1981). Thus, phosphorylation of N_{TAIL} might modulate the structural state of N. In agreement, MV N° and N^{NUC} are phosphorylated with a different phosphorylation pattern (Gombart *et al.*, 1995).

CONCLUSION

It is likely that N and P contain large regions that are unstructured even in complex. Indeed, the whole region 145-320 of SeV P is dispensable for transcription and replication (Curran *et al.*, 1995b). Even assuming that the region 1-144 of SeV P folds upon binding to N°, it is likely that aa 145-320 are mainly unstructured in the N°-P complex. Therefore, once the structure of modular, structured domains of N and P has been solved, structural advances in the study of the replication complex of *Mononegavirales* will either i) need removal of several dispensable, unstructured regions for crystallization of protein complexes, a trial-and-error process likely to be very time-consuming; ii) rely on techniques that can deal with both ordered and disordered regions of proteins (such as small angle X-ray scattering); iii) rely on "minimalist" viruses which have less dispensable, unstructured regions, such as *Pneumovirinae*; iv) rely on the comparative study of viruses by biocomputing methods to identify further functional regions in poorly conserved, unstructured parts of N,P, and L.
From a fundamental point of view, we think that the replication complex of *Mononegavirales* could constitute a model system for the emerging field of study of structural disorder in proteins.

Acknowledgements

We wish to thank F. Ferron for help with the multiple sequence editor Seaview and K. Dunker for making PONDR available to us. Thanks to J. Curran, D. Gerlier, and L. Roux for critical reading of the manuscript. Access to PONDR was provided by Molecular Kinetics (P.O. Box 2475 CS, Pullman, WA 99165-2475; E-mail: mkinetics@turbonet.com) under license from the WSU Research Foundation. PONDR is copyright ©1999 by the WSU Research Foundation, all rights reserved. This work was supported by a grant from the Fondation pour la Recherche Médicale (FRM) to D. Karlin. This study has been carried out with financial support from the Commission of the European Communities, specific RTD programme "Quality of Life and Management of Living Resources", QLK2-CT2001-01225, "Towards the design of new potent antiviral drugs: structure-function analysis of Paramyxoviridae RNA polymerase". It does not necessarily reflect its views and in no way anticipates the Commission's future policy in this area.

REFERENCES

Baker, S. C., and Moyer, S. A. (1988). Encapsidation of Sendai virus genome RNAs by purified NP protein during in vitro replication. *J Virol* **62**(3)**,** 834-8.

Bankamp, B., Horikami, S. M., Thompson, P. D., Huber, M., Billeter, M., and Moyer, S. A. (1996). Domains of the measles virus N protein required for binding to P protein and self-assembly. *Virology* **216**(1)**,** 272-7.

Bohn, W., Ciampor, F., Rutter, R., and Mannweiler, K. (1990). Localization of nucleocapsid associated polypeptides in measles virus- infected cells by immunogold labelling after resin embedding. *Arch Virol* **114**(1-2)**,** 53-64.

Bousse, T., Takimoto, T., Matrosovich, T., and Portner, A. (2001). Two regions of the P protein are required to be active with the L protein for human parainfluenza virus type 1 RNA polymerase activity. *Virology* **283**(2)**,** 306-14.

Bowman, M. C., Smallwood, S., and Moyer, S. A. (1999). Dissection of individual functions of the Sendai virus phosphoprotein in transcription. *J. Virol.* **73**(8)**,** 6474-83.

Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T., Oldfield, C. J., Williams, C. J., and Dunker, A. K. (2002). Titre non communiqué. *J. Mol. Evol.* ***in press.***

Buchholz, C. J., Retzler, C., Homann, H. E., and Neubert, W. J. (1994). The carboxy-terminal domain of Sendai virus nucleocapsid protein is involved in complex formation between phosphoprotein and nucleocapsid- like particles. *Virology* **204**(2)**,** 770-6.

Buchholz, C. J., Spehner, D., Drillien, R., Neubert, W. J., and Homann, H. E. (1993). The conserved N-terminal region of Sendai virus nucleocapsid protein NP is required for nucleocapsid assembly. *J Virol* **67**(10)**,** 5803-12.

Byrappa, S., and Gupta, K. C. (1999). Human parainfluenza virus type 1 phosphoprotein is constitutively phosphorylated at Ser-120 and Ser-184. *J. Gen. Virol.* **80**(Pt 5)**,** 1199-209.

Byrappa, S., Pan, Y. B., and Gupta, K. C. (1996). Sendai virus P protein is constitutively phosphorylated at serine249: high phosphorylation potential of the P protein. *Virology* **216**(1)**,** 228-34.

Cadd, T., Garcin, D., Tapparel, C., Itoh, M., Homma, M., Roux, L., Curran, J., and Kolakofsky, D. (1996). The Sendai paramyxovirus accessory C proteins inhibit viral genome amplification in a promoter-specific fashion. *J. Virol.* **70**(8)**,** 5067-74.

Chandrika, R., Horikami, S. M., Smallwood, S., and Moyer, S. A. (1995). Mutations in conserved domain I of the Sendai virus L polymerase protein uncouple transcription and replication. *Virology* **213**(2)**,** 352-63.

Chinchar, V. G., and Portner, A. (1981a). Functions of Sendai virus nucleocapsid polypeptides: enzymatic activities in nucleocapsids following cleavage of polypeptide P by Staphylococcus aureus protease V8. *Virology* **109**(1)**,** 59-71.

Chinchar, V. G., and Portner, A. (1981b). Inhibition of RNA synthesis following proteolytic cleavage of Newcastle disease virus P protein. *Virology* **115**(1)**,** 192-202.

Choudhary, S., De, B. P., and Banerjee, A. K. (2000). Specific phosphorylated forms of glyceraldehyde 3-phosphate dehydrogenase associate with human parainfluenza virus type 3 and inhibit viral transcription in vitro. *J Virol* **74**(8)**,** 3634-41.

Coronel, E. C., Takimoto, T., Murti, K. G., Varich, N., and Portner, A. (2001). Nucleocapsid incorporation into parainfluenza virus is regulated by specific interaction with matrix protein. *J Virol* **75**(3)**,** 1117-23.

Cortese, C. K., Feller, J. A., and Moyer, S. A. (2000). Mutations in domain V of the Sendai virus L polymerase protein uncouple transcription and replication and differentially affect replication in vitro and in vivo. *Virology* **277**(2)**,** 387-96.

Curran, J. (1996). Reexamination of the Sendai virus P protein domains required for RNA synthesis: a possible supplemental role for the P protein. *Virology* **221**(1)**,** 130-40.

Curran, J. (1998). A role for the Sendai virus P protein trimer in RNA synthesis. *J. Virol.* **72**(5)**,** 4274-80.

Curran, J., Boeck, R., and Kolakofsky, D. (1991). The Sendai virus P gene expresses both an essential protein and an inhibitor of RNA synthesis by shuffling modules via mRNA editing. *EMBO J.* **10**(10)**,** 3079-85.

Curran, J., Boeck, R., Lin-Marq, N., Lupas, A., and Kolakofsky, D. (1995a). Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils. *Virology* **214**(1)**,** 139-49.

Curran, J., Homann, H., Buchholz, C., Rochat, S., Neubert, W., and Kolakofsky, D. (1993). The hypervariable C-terminal tail of the Sendai paramyxovirus nucleocapsid protein is required for template function but not for RNA encapsidation. *J. Virol.* **67**(7)**,** 4358-64.

Curran, J., and Kolakofsky, D. (1999). Replication of paramyxoviruses. *Adv. Virus Res.* **54,** 403-22.

Curran, J., Marq, J. B., and Kolakofsky, D. (1992). The Sendai virus nonstructural C proteins specifically inhibit viral mRNA synthesis. *Virology* **189**(2)**,** 647-56.

Curran, J., Marq, J. B., and Kolakofsky, D. (1995b). An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J. Virol.* **69**(2)**,** 849-55.

Curran, J., Pelet, T., and Kolakofsky, D. (1994). An acidic activation-like domain of the Sendai virus P protein is required for RNA synthesis and encapsidation. *Virology* **202**(2)**,** 875-84.

Das, T., Schuster, A., Schneider-Schaulies, S., and Banerjee, A. K. (1995). Involvement of cellular casein kinase II in the phosphorylation of measles virus P protein: identification of phosphorylation sites. *Virology* **211**(1)**,** 218-26.

De, B. P., and Banerjee, A. K. (1999). Involvement of actin microfilaments in the transcription/replication of human parainfluenza virus type 3: possible role of actin in other viruses. *Microsc Res Tech* **47**(2)**,** 114-23.

De, B. P., Das, T., and Banerjee, A. K. (1997). Role of cellular kinases in the gene expression of nonsegmented negative strand RNA viruses. *Biol. Chem.* **378**(6)**,** 489-93.

De, B. P., Hoffman, M. A., Choudhary, S., Huntley, C. C., and Banerjee, A. K. (2000). Role of NH(2)- and COOH-terminal domains of the P protein of human parainfluenza virus type 3 in transcription and replication. *J. Virol.* **74**(13)**,** 5886-95.

Deshpande, K. L., and Portner, A. (1984). Structural and functional analysis of Sendai virus nucleocapsid protein NP with monoclonal antibodies. *Virology* **139**(1)**,** 32-42.

Deshpande, K. L., and Portner, A. (1985). Monoclonal antibodies to the P protein of Sendai virus define its structure and role in transcription. *Virology* **140**(1)**,** 125-34.

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001). Intrinsically disordered protein. *J. Mol. Graph. Model.* **19**(1)**,** 26-59.

Dyson, H. J., and Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **12**(1)**,** 54-60.

Egelman, E. H., Wu, S. S., Amrein, M., Portner, A., and Murti, G. (1989). The Sendai virus nucleocapsid exists in at least four different helical states. *J Virol* **63**(5)**,** 2233-43.

Errington, W., and Emmerson, P. T. (1997). Assembly of recombinant Newcastle disease virus nucleocapsid protein into nucleocapsid-like structures is inhibited by the phosphoprotein. *J Gen Virol* **78**(Pt 9)**,** 2335-9.

Feller, J. A., Smallwood, S., Horikami, S. M., and Moyer, S. A. (2000). Mutations in conserved domains IV and VI of the large (L) subunit of the sendai virus RNA polymerase give a spectrum of defective RNA synthesis phenotypes. *Virology* **269**(2)**,** 426-39.

Ferron, F., Longhi, S., Henrissat, B., and Canard, B. (2002). Viral RNA-polymerases - a predicted 2-'O ribose methyltransferase domain shared by all *Mononegavirales*. *TIBS* **27**(5)**,** 222-4.

Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1997). Probing the partly folded states of proteins by limited proteolysis. *Fold. Des.* **2**(2)**,** R17-26.

Gaudier, M., Gaudin, Y., and Knossow, M. (2001). Cleavage of vesicular stomatitis virus matrix protein prevents self- association and leads to crystallization. *Virology* **288**(2)**,** 308-14.

Gombart, A. F., Hirano, A., and Wong, T. C. (1993). Conformational maturation of measles virus nucleocapsid protein. *J. Virol.* **67**(7)**,** 4133-41.

Gombart, A. F., Hirano, A., and Wong, T. C. (1995). Nucleoprotein phosphorylated on both serine and threonine is preferentially assembled into the nucleocapsids of measles virus. *Virus Res* **37**(1)**,** 63-73.

Griffin, D. E. (2001). Measles Virus. ***In*** "Fields Virology" (B.N. Fields, D.M. Knipe, and P.M. Howley, Eds.), 4th ed., pp 1401-1441 Lippincott-Raven, Philadelphia, PA.

Grogan, C. C., and Moyer, S. A. (2001). Sendai virus wild-type and mutant C proteins show a direct correlation between L polymerase binding and inhibition of viral RNA synthesis. *Virology* **288**(1)**,** 96-108.

Gupta, A. K., and Banerjee, A. K. (1997). Expression and purification of vesicular stomatitis virus N-P complex from Escherichia coli: role in genome RNA transcription and replication in vitro. *J Virol* **71**(6)**,** 4264-71.

Harty, R. N., and Palese, P. (1995). Measles virus phosphoprotein (P) requires the NH2- and COOH-terminal domains for interactions with the nucleoprotein (N) but only the COOH terminus for interactions with itself. *J. Gen. Virol.* **76**(Pt 11)**,** 2863-7.

Heggeness, M. H., Scheid, A., and Choppin, P. W. (1980). Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration. *Proc Natl Acad Sci U S A* **77**(5)**,** 2631-5.

Heggeness, M. H., Scheid, A., and Choppin, P. W. (1981). The relationship of conformational changes in the Sendai virus nucleocapsid to proteolytic cleavage of the NP polypeptide. *Virology* **114**(2)**,** 555-62.

Holmes, D. E., and Moyer, S. A. (2002). The phosphoprotein (p) binding site resides in the N terminus of the L polymerase subunit of sendai virus. *J Virol* **76**(6)**,** 3078-83.

Horikami, S. M., Curran, J., Kolakofsky, D., and Moyer, S. A. (1992). Complexes of Sendai virus NP-P and P-L proteins are required for defective interfering particle genome replication in vitro. *J. Virol.* **66**(8)**,** 4901-8.

Horikami, S. M., Hector, R. E., Smallwood, S., and Moyer, S. A. (1997). The Sendai virus C protein binds the L polymerase protein to inhibit viral RNA synthesis. *Virology* **235**(2)**,** 261-70.

Horikami, S. M., and Moyer, S. A. (1995). Alternative amino acids at a single site in the Sendai virus L protein produce multiple defects in RNA synthesis in vitro. *Virology* **211**(2)**,** 577-82.

Horikami, S. M., Smallwood, S., Bankamp, B., and Moyer, S. A. (1994). An amino-proximal domain of the L protein binds to the P protein in the measles virus RNA polymerase complex. *Virology* **205**(2)**,** 540-5.

Horikami, S. M., Smallwood, S., and Moyer, S. A. (1996). The Sendai virus V protein interacts with the NP protein to regulate viral genome RNA replication. *Virology* **222**(2)**,** 383-90.

Hsu, C. H., and Kingsbury, D. W. (1982). Topography of phosphate residues in Sendai virus proteins. *Virology* **120**(1)**,** 225-34.

Hu, C., and Gupta, K. C. (2000). Functional significance of alternate phosphorylation in Sendai virus P protein. *Virology* **268**(2)**,** 517-32.

Hu, C. J., Kato, A., Bowman, M. C., Kiyotani, K., Yoshida, T., Moyer, S. A., Nagai, Y., and Gupta, K. C. (1999). Role of primary constitutive phosphorylation of Sendai virus P and V proteins in viral replication and pathogenesis. *Virology* **263**(1)**,** 195-208.

Huang, C., Kiyotani, K., Fujii, Y., Fukuhara, N., Kato, A., Nagai, Y., Yoshida, T., and Sakaguchi, T. (2000). Involvement of the zinc-binding capacity of Sendai virus V protein in viral pathogenesis. *J Virol* **74**(17)**,** 7834-41.

Hubbard, S. J. (1998). The structural aspects of limited proteolysis of native proteins. *Biochim Biophys Acta* **1382**(2)**,** 191-206.

Hubbard, S. J., Beynon, R.J. (2001). Proteolysis of native proteins as a structural probe. *In " Proteolytic Enzymes" (Beynon R.J., Bond J.D., Eds.), 2nd ed., pp 233-264. Oxford University Press, Oxford, NY.*

Huber, M., Cattaneo, R., Spielhofer, P., Orvell, C., Norrby, E., Messerli, M., Perriard, J. C., and Billeter, M. A. (1991). Measles virus phosphoprotein retains the nucleocapsid protein in the cytoplasm. *Virology* **185**(1)**,** 299-308.

Huntley, C. C., De, B. P., and Banerjee, A. K. (1997). Phosphorylation of Sendai virus phosphoprotein by cellular protein kinase C zeta. *J. Biol. Chem.* **272**(26)**,** 16578-84.

Iseni, F., Baudin, F., Blondel, D., and Ruigrok, R. W. (2000). Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. *Rna* **6**(2)**,** 270-81.

Jonscher, K. R., and Yates, J. R., 3rd (1997). Matrix-assisted laser desorption ionization/quadrupole ion trap mass spectrometry of peptides. Application to the localization of phosphorylation sites on the P protein from Sendai virus. *J Biol Chem* **272**(3)**,** 1735-41.

Jordan, I. K., Sutter, B. A. t., and McClure, M. A. (2000). Molecular evolution of the Paramyxoviridae and Rhabdoviridae multiple-protein-encoding P gene. *Mol Biol Evol* **17**(1)**,** 75-86.

Karlin, D., Longhi, S., and Canard, B. (2002a). Substitution of two residues in the measles virus nucleoprotein results in an impaired self-association. **Submitted for publication**.

Karlin, D., Longhi, S., Receveur, V., and Canard, B. (2002b). The measles virus phosphoprotein N-terminal domain (PNT) belongs to the natively unfolded class of proteins. *Virology* **In press**.

Khattar, S. K., Yunus, A. S., Collins, P. L., and Samal, S. K. (2000). Mutational analysis of the bovine respiratory syncytial virus nucleocapsid protein using a minigenome system: mutations that affect encapsidation, RNA synthesis, and interaction with the phosphoprotein. *Virology* **270**(1)**,** 215-28.

Kouznetzoff, A., Buckle, M., and Tordo, N. (1998). Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. *J Gen Virol* **79**(Pt 5)**,** 1005-13.

Lai, M. M. (1998). Cellular factors in the transcription and replication of viral RNA genomes: a parallel to DNA-dependent RNA transcription. *Virology* **244**(1)**,** 1-12.

Lamb, R. A., and Choppin, P. W. (1977). The synthesis of Sendai virus polypeptides in infected cells. II. Intracellular distribution of polypeptides. *Virology* **81**(2)**,** 371-81.

Lamb, R. A., and Kolakofsky, D. (2001). *Paramyxoviridae* : The Viruses and Their Replication. ***In*** "Fields Virology" (B.N. Fields, D.M. Knipe, and P.M. Howley, Eds.), 4th ed., pp 1305-1340. Lippincott-Raven, Philadelphia, PA.

Latorre, P., Kolakofsky, D., and Curran, J. (1998). Sendai virus Y proteins are initiated by a ribosomal shunt. *Mol Cell Biol* **18**(9)**,** 5021-31.

Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999). Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform. Ser. Workshop Genome Inform.* **10,** 30-40.

Lin, G. Y., Paterson, R. G., Richardson, C. D., and Lamb, R. A. (1998). The V protein of the paramyxovirus SV5 interacts with damage-specific DNA binding protein. *Virology* **249**(1)**,** 189-200.

Liston, P., Batal, R., DiFlumeri, C., and Briedis, D. J. (1997). Protein interaction domains of the measles virus nucleocapsid protein (NP). *Arch Virol* **142**(2)**,** 305-21.

Liston, P., DiFlumeri, C., and Briedis, D. J. (1995). Protein interactions entered into by the measles virus P, V, and C proteins. *Virus Res.* **38**(2-3)**,** 241-59.

Liu, Z., Huntley, C. C., De, B. P., Das, T., Banerjee, A. K., and Oglesbee, M. J. (1997). Phosphorylation of canine distemper virus P protein by protein kinase C- zeta and casein kinase II. *Virology* **232**(1)**,** 198-206.

Marie, J. C., Kehren, J., Trescol-Biemont, M. C., Evlashev, A., Valentin, H., Walzer, T., Tedone, R., Loveland, B., Nicolas, J. F., Rabourdin-Combe, C., and Horvat, B. (2001). Mechanism of measles virus-induced suppression of inflammatory immune responses. *Immunity* **14**(1)**,** 69-79.

Marion, D., Tarbouriech, N., Ruigrok, R. W., Burmeister, W. P., and Blanchard, L. (2001). Assignment of the 1H, 15N and 13C resonances of the nucleocapsid- binding domain of the Sendai virus phosphoprotein. *J Biomol NMR* **21**(1)**,** 75-6.

Melcher, K. (2000). The strength of acidic activation domains correlates with their affinity for both transcriptional and non-transcriptional proteins. *J. Mol. Biol.* **301**(5)**,** 1097-112.

Murphy, S. K., Ito, Y., and Parks, G. D. (1998). A functional antigenomic promoter for the paramyxovirus simian virus 5 requires proper spacing between an essential internal segment and the 3' terminus. *J Virol* **72**(1)**,** 10-9.

Murphy, S. K., and Parks, G. D. (1999). RNA replication for the paramyxovirus simian virus 5 requires an internal repeated (CGNNNN) sequence motif. *J Virol* **73**(1)**,** 805-9.

Murray, J., Loney, C., Murphy, L. B., Graham, S., and Yeo, R. P. (2001). Characterization of Monoclonal Antibodies Raised against Recombinant Respiratory Syncytial Virus Nucleocapsid (N) Protein: Identification of a Region in the Carboxy Terminus of N Involved in the Interaction with P Protein. *Virology* **289**(2)**,** 252-61.

Myers, T. M., and Moyer, S. A. (1997). An amino-terminal domain of the Sendai virus nucleocapsid protein is required for template function in viral RNA synthesis. *J. Virol.* **71**(2)**,** 918-24.

Myers, T. M., Pieters, A., and Moyer, S. A. (1997). A highly conserved region of the Sendai virus nucleocapsid protein contributes to the NP-NP binding domain. *Virology* **229**(2)**,** 322-35.

Myers, T. M., Smallwood, S., and Moyer, S. A. (1999). Identification of nucleocapsid protein residues required for Sendai virus nucleocapsid formation and genome replication. *J. Gen. Virol.* **80**(Pt 6)**,** 1383-91.

Nishio, M., Tsurudome, M., Ito, M., and Ito, Y. (2000). Mapping of domains on the human parainfluenza type 2 virus P and NP proteins that are involved in the interaction with the L protein. *Virology* **273**(2)**,** 241-7.

Nishio, M., Tsurudome, M., Ito, M., Kawano, M., Kusagawa, S., Komada, H., and Ito, Y. (1999). Mapping of domains on the human parainfluenza virus type 2 nucleocapsid protein (NP) required for NP-phosphoprotein or NP-NP interaction. *J Gen Virol* **80**(Pt 8)**,** 2017-22.

Nishio, M., Tsurudome, M., Ito, M., Watanabe, N., Kawano, M., Komada, H., and Ito, Y. (1997). Human parainfluenza virus type 2 phosphoprotein: mapping of monoclonal antibody epitopes and location of the multimerization domain. *J. Gen. Virol.* **78**(Pt 6)**,** 1303-8.

Nishio, M., Tsurudome, M., Kawano, M., Watanabe, N., Ohgimoto, S., Ito, M., Komada, H., and Ito, Y. (1996). Interaction between nucleocapsid protein (NP) and phosphoprotein (P) of human parainfluenza virus type 2: one of the two NP binding sites on P is essential for granule formation. *J. Gen. Virol.* **77**(Pt 10)**,** 2457-63.

Parks, G. D. (1994). Mapping of a region of the paramyxovirus L protein required for the formation of a stable complex with the viral phosphoprotein P. *J. Virol.* **68**(8)**,** 4862-72.

Paterson, R. G., Leser, G. P., Shaughnessy, M. A., and Lamb, R. A. (1995). The paramyxovirus SV5 V protein binds two atoms of zinc and is a structural component of virions. *Virology* **208**(1)**,** 121-31.

Poch, O., Blumberg, B. M., Bougueleret, L., and Tordo, N. (1990). Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *J Gen Virol* **71**(Pt 5)**,** 1153-62.

Portner, A., and Murti, K. G. (1986). Localization of P, NP, and M proteins on Sendai virus nucleocapsid using immunogold labeling. *Virology* **150**(2)**,** 469-78.

Portner, A., Murti, K. G., Morgan, E. M., and Kingsbury, D. W. (1988). Antibodies against Sendai virus L protein: distribution of the protein in nucleocapsids revealed by immunoelectron microscopy. *Virology* **163**(1)**,** 236-9.

Precious, B., Young, D. F., Bermingham, A., Fearns, R., Ryan, M., and Randall, R. E. (1995). Inducible expression of the P, V, and NP genes of the paramyxovirus simian virus 5 in cell lines and an examination of NP-P and NP-V interactions. *J. Virol.* **69**(12)**,** 8001-10.

Randall, R. E., and Bermingham, A. (1996). NP:P and NP:V interactions of the paramyxovirus simian virus 5 examined using a novel protein:protein capture assay. *Virology* **224**(1)**,** 121-9.

Ravanel, K., Castelle, C., Defrance, T., Wild, T. F., Charron, D., Lotteau, V., and Rabourdin-Combe, C. (1997). Measles virus nucleocapsid protein binds to FcgammaRII and inhibits human B cell antibody production. *J Exp Med* **186**(2)**,** 269-78.

Reutter, G. L., Cortese-Grogan, C., Wilson, J., and Moyer, S. A. (2001). Mutations in the measles virus C protein that up regulate viral RNA synthesis. *Virology* **285**(1)**,** 100-9.

Romero, P., Obradovic, Z., and Dunker, A. K. (1999). Folding minimal sequences: the lower bound for sequence complexity of globular proteins. *FEBS Lett.* **462**(3)**,** 363-7.

Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., and Dunker, A. K. (1997). Identifying disordered regions in proteins from amino acid sequences. *Proc. I.E.E.E. International Conference on*

Neural Networks **1,** 90-95.

Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., and Dunker, A. K. (1998). Thousands of proteins likely to have long disordered regions. ***In*** "Pacific Symposium on Biocomputing 98", (R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, Eds.), pp 437-48. World Scientific Pub Co, London.

Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001). Sequence complexity of disordered proteins. *Proteins* **42**(1)**,** 38-48.

Rost, B. (1996). PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol* **266,** 525-39.

Ryan, K. W., Morgan, E. M., and Portner, A. (1991). Two noncontiguous regions of Sendai virus P protein combine to form a single nucleocapsid binding domain. *Virology* **180**(1)**,** 126-34.

Ryan, K. W., Murti, K. G., and Portner, A. (1990). Localization of P protein binding sites on the Sendai virus nucleocapsid. *J. Gen. Virol.* **71**(Pt 4)**,** 997-1000.

Ryan, K. W., and Portner, A. (1990). Separate domains of Sendai virus P protein are required for binding to viral nucleocapsids. *Virology* **174**(2)**,** 515-21.

Ryan, K. W., Portner, A., and Murti, K. G. (1993). Antibodies to paramyxovirus nucleoproteins define regions important for immunogenicity and nucleocapsid assembly. *Virology* **193**(1)**,** 376-84.

Schoehn, G., Iseni, F., Mavrakis, M., Blondel, D., and Ruigrok, R. W. (2001). Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. *J. Virol.* **75**(1)**,** 490-8.

Sedlmeier, R., and Neubert, W. J. (1998). The replicative complex of paramyxoviruses: structure and function. *Adv. Virus Res.* **50,** 101-39.

Shaji, D., and Shaila, M. S. (1999). Domains of Rinderpest virus phosphoprotein involved in interaction with itself and the nucleocapsid protein. *Virology* **258**(2)**,** 415-24.

Sidhu, M. S., Menonna, J. P., Cook, S. D., Dowling, P. C., and Udem, S. A. (1993). Canine distemper virus L gene: sequence and comparison with related viruses. *Virology* **193**(1)**,** 50-65.

Smallwood, S., Easson, C. D., Feller, J. A., Horikami, S. M., and Moyer, S. A. (1999). Mutations in conserved domain II of the large (L) subunit of the Sendai virus RNA polymerase abolish RNA synthesis. *Virology* **262**(2)**,** 375-83.

Smallwood, S., Ryan, K. W., and Moyer, S. A. (1994). Deletion analysis defines a carboxyl-proximal region of Sendai virus P protein that binds to the polymerase L protein. *Virology* **202**(1)**,** 154-63.

Spehner, D., Drillien, R., and Howley, P. M. (1997). The assembly of the measles virus nucleoprotein into nucleocapsid-like particles is modulated by the phosphoprotein. *Virology* **232**(2)**,** 260-8.

Svenda, M., Hjertner, B., Linne, T., and Berg, M. (2002). Both the P and V proteins of the porcine rubulavirus LPMV interact with the NP protein via their respective C-terminal unique parts. *Virus Res* **83**(1-2)**,** 31-41.

Sweetman, D. A., Miskin, J., and Baron, M. D. (2001). Rinderpest virus C and V proteins interact with the major (L) component of the viral polymerase. *Virology* **281**(2)**,** 193-204.

Tapparel, C., Maurice, D., and Roux, L. (1998). The activity of Sendai virus genomic and antigenomic promoters requires a second element past the leader template regions: a motif (GNNNNN)3 is essential for replication. *J Virol* **72**(4)**,** 3117-28.

Tarbouriech, N., Curran, J., Ruigrok, R. W., and Burmeister, W. P. (2000). Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nat. Struct. Biol.* **7**(9)**,** 777-81.

Tober, C., Seufert, M., Schneider, H., Billeter, M. A., Johnston, I. C., Niewiesk, S., ter Meulen, V., and Schneider-Schaulies, S. (1998). Expression of measles virus V protein is associated with pathogenicity and control of viral RNA synthesis. *J. Virol.* **72**(10)**,** 8124-32.

Tuckis, J., Smallwood, S., Feller, J. A., and Moyer, S. A. (2002). The C-terminal 88 amino acids of the Sendai virus P protein have multiple functions separable by mutation. *J Virol* **76**(1)**,** 68-77.

Uversky, V. N. (2002). What does it mean to be natively unfolded? *Eur J Biochem* **269**(1)**,** 2-12.

Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41**(3)**,** 415-427.

Vidal, S., Curran, J., Orvell, C., and Kolakofsky, D. (1988). Mapping of monoclonal antibodies to the Sendai virus P protein and the location of its phosphates. *J Virol* **62**(6)**,** 2200-3.

Wang, L. F., Yu, M., Hansson, E., Pritchard, L. I., Shiell, B., Michalski, W. P., and Eaton, B. T. (2000). The exceptionally large genome of Hendra virus: support for creation of a new genus within the family Paramyxoviridae. *J Virol* **74**(21)**,** 9972-9.

Warnes, A., Fooks, A. R., Dowsett, A. B., Wilkinson, G. W., and Stephenson, J. R. (1995). Expression of the measles virus nucleoprotein gene in Escherichia coli and assembly of nucleocapsid-like structures. *Gene* **160**(2)**,** 173-8.

Watanabe, N., Kawano, M., Tsurudome, M., Nishio, M., Ito, M., Ohgimoto, S., Suga, S., Komada, H., and Ito, Y. (1996). Binding of the V proteins to the nucleocapsid proteins of human parainfluenza type 2 virus. *Med. Microbiol. Immunol. (Berl)* **185**(2)**,** 89-94.

Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., and Lansbury, P. T., Jr. (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* **35**(43)**,** 13709-15.

Williams, R. M., Obradovi, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., Takayama, S., Brown, C. J., and Dunker, A. K. (2001). The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac. Symp. Biocomput.***,** 89-100.

Wright, P. E., and Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**(2)**,** 321-31.

Zetina, C. R. (2001). A conserved helix-unfolding motif in the naturally unfolded proteins. *Proteins* **44**(4)**,** 479-83.

Zhao, H., and Banerjee, A. K. (1995). Interaction between the nucleocapsid protein and the phosphoprotein of human parainfluenza virus 3. Mapping of the interacting domains using a two-hybrid system. *J Biol Chem* **270**(21)**,** 12485-90.

Alors que nous pouvons nous représenter clairement la *structure* des protéines, pour l'instant nous ne disposons pas d'outils conceptuels adéquats pour nous représenter les régions désordonnées des protéines. Nous ne savons d'ailleurs pas si une telle représentation est possible, ou aurait un sens. Pour prendre un exemple, dans un moteur, les pièces solides (engrenages, par exemple) ont une forme distincte et constante, et leur fonction dépend de cette forme. Les pièces souples (courroie, par exemple), peuvent au contraire adopter une multitude de formes, et doivent faire preuve de qualités différentes de celles des engrenages. Cependant, dans un moteur, une courroie adopte une forme distincte, qui est nécessaire à sa fonction. Dans ce cas précis, ce n'est pas le *désordre* structural de la courroie qui est intéressant, ce sont ses propriétés de *souplesse*, *d'adhérence* et de *résistance* - qui sont liées au matériau choisi. L'absence de forme prédéfinie n'est qu'un corollaire des propriétés de ce matériau, mais ce n'est pas elle qui fait l'importance de la courroie dans le moteur. De la même façon, peut-être sont-ce les propriétés de souplesse, d'élasticité, ou de résistance de certaines protéines partiellement désordonnées qui sont intéressantes (à l'exemple de la titine, une protéine géante des fibres musculaires) ; ces propriétés dérivent d'un arrangement particulier d'acides aminés (le matériau) … Au contraire, ce pourrait être la sensibilité de certaines protéines désordonnées à la dégradation protéolytique qui fait tout leur intérêt. Notons toutefois que dans une cellule, beaucoup de protéines désordonnées sont parfaitement stables, ce qui suggère, ou bien qu'elles sont physiquement séparées des protéases, ou encore qu'elles adoptent *in vivo* une conformation résistante à la protéolyse (par exemple lors d'une association avec un partenaire).

Les *Morbillivirus*, un système modèle pour l'étude du désordre structural ?

Les régions désordonnées de protéines modulaires homologues sont souvent de longueur très variable et sans homologie de séquence, en particulier quand elles jouent le rôle d'un élément de liaison ("linker") flexible (B. Henrissat, communication personnelle). Toutefois, il est possible d'établir un alignement multiple de très bonne qualité de deux régions non structurées de protéines des *Morbillivirus* (PNT et N_{TAIL}), malgré leur variabilité en séquence. Ce n'est pas le cas pour PNT chez les *Respirovirus* ou les *Rubulavirus*, dont les séquences sont trop variables et/ou de longueur différentes).

Les régions PNT et N_{TAIL} des *Morbillivirus* sont donc susceptibles de constituer un modèle idéal pour étudier les propriétés de séquence de protéines désordonnées, et identifier des caractères importants conservés à travers une famille de protéines homologues. En effet :

l'identification dans une région de certains *attributs de séquence conservés*, alors même que la *séquence* en acides aminés est *variable*, suggèrerait une importance fonctionnelle de cette région.

Corréler de telles indications à des informations fonctionnelles demande bien sûr de connaître les régions fonctionnelles de PNT et de N_{TAIL} de façon plus précise qu'actuellement. Cette cartographie est difficile à établir précisément en l'absence de toute structure tridimensionnelle pour la guider. Cependant, il existe déjà un répertoire intéressant de mutations (nombreuses) dans ces protéines, dont on sait qu'elles sont impliquées dans l'atténuation virale (à titre d'exemple: (Parks *et coll.*, 2001; Takeda *et coll.*, 1998)), ce qui peut simplifier la cartographie.

Bien entendu, *à contrario,* l'analyse bio-informatiques peut fournir des pistes quant aux régions fonctionnelles, suppléant ainsi à une structure tridimensionnelle impossible à obtenir. À ce titre, je pense que le programme REAL (C. Abergel, Marseille, non publié), qui permet de mettre en regard des régions variables avec la présence d'un paramètre invariant (hydrophobicité, pourcentage en tel acide aminé, etc.) serait un outil idéal. Cette étude sera poursuivie après ma thèse, en collaboration avec K. Dunker (Portland, USA) (pour un exemple d'une étude de l'évolution comparée des régions désordonnées : (Brown *et coll.*, 2002). Cet auteur est particulièrement intéressé par les virus à ARN en raison de leur évolution très rapide, due à une faible fidélité des ARN-polymérases.

Sur un plan expérimental et non plus bio-informatique, notons la disponibilité récente au laboratoire d'une technique extrêmement puissante d'analyse des protéines ou des complexes protéiques, la diffusion des rayons X aux petits angles (SAXS). Elle est particulièrement adaptée à l'étude des protéines contenant un mélange de régions désordonnées et ordonnées (pour exemple : (Grossmann *et coll.*, 2001)). Son emploi pourra faire énormément progresser la connaissance de complexes tels que N°-P et N^{NUC}-P. Gageons que le complexe de réplication viral des *Paramyxoviridae* pourra alors devenir un système modèle pour l'étude biochimique et biophysique du désordre structural …

PERSPECTIVES – LE COMPLEXE DE REPLICATION VIRAL : PROBLEMATIQUES STRUCTURALES

Les travaux présentés dans les deux articles et dans la revue soulèvent un certain nombre de questions. Dans ce chapitre, nous présentons certaines d'entre elles comme des questions ouvertes ; pour d'autres, nous proposons des pistes de recherches.

Maturation conformationnelle de la nucléoprotéine

La figure 3 de la revue illustre une interprétation originale de la maturation conformationnelle de la nucléoprotéine du virus de la rougeole, fondée sur l'article (Gombart *et coll.*, 1993) mais aussi sur des articles postérieurs comme (Spehner *et coll.*, 1997).
Les travaux de Baker *et coll.* suggèrent que le complexe d'assemblage N°-P est nécessaire uniquement à **l'initiation** de la réplication (Baker and Moyer, 1988). Une forme soluble de N (N° ?) suffirait à encapsider l'ARN néosynthétisé durant son **élongation**. Ce modèle est assez séduisant, car il y a effectivement une différence majeure entre l'étape d'initiation et celle d'élongation. En effet, lors de l'initiation de la polymérisation de N sur l'ARN, une (ou plusieurs ?) molécule de N° doit adopter (en l'absence de molécule de N^{NUC}) une conformation qui lui permette de se lier à l'ARN néosynthétisé. En revanche, toutes les étapes postérieures d'ajout de sous-unités au polymère de N sont identiques : N° se lie à l'ARN néosynthétisé *et* aux molécules de N^{NUC} déjà polymérisées, ce qui doit grandement faciliter son changement de conformation.

Nous avons les moyens au laboratoire d'étudier les différents complexes entre N et P. Des problématiques telles que le changement de conformation de la nucléocapside en bas sel (Heggeness *et coll.*, 1980; Heggeness *et coll.*, 1981) ainsi que l'étude d'un changement éventuel de conformation de la nucléocapside induite par P seront poursuivies au laboratoire.

Analyse de l'organisation fonctionnelle et structurale de N à travers les *Paramyxoviridae*

Je pense que nous disposons de suffisamment d'éléments pour essayer d'établir par des méthodes bio-informatiques une cartographie de l'organisation de N complémentaire de celle que peuvent nous offrir les études de délétion. En effet, N_{CORE} n'ayant pas une structure modulaire, celles-ci se sont principalement limitées à cartographier les régions d'interaction entre N^{NUC}_{TAIL}, $N°_{TAIL}$, P et L. Les séquences de nucléoprotéines des *Paramyxoviridae* sont en nombre de plus en plus important (plus de 20 à ce jour). Ce travail de séquençage et de découverte de nouveaux virus revêt une grande importance. En effet :

Certaines régions de N sont extrêmement bien conservées dans les *Paramyxovirinae*, alors que d'autres régions ne sont conservées que dans un genre viral. En mettant en parallèle ces informations avec les indications de régions conservées entre différents genres viraux dans la protéine P, il est permis de penser qu'on devrait pouvoir délimiter les régions probables d'interaction entre N° et P, et/ou entre N et N, mais aussi d'autres régions fonctionnelles.

C'est le principe de l'étude de **l'évolution corrélée de régions protéiques**, encore appelé **exploitation du contexte des régions protéiques**, schématisé en figure 8. Une fois de telles régions délimitées, leur validation, ainsi qu'une analyse plus fine au niveau des acides aminés, devrait être possible en utilisant les logiciels d'analyse de substitutions corrélées CRASP (Afonnikov *et coll.*, 2001), ou Plotcorr (Pazos *et coll.*, 1997a; Pazos *et coll.*, 1997b).

Figure 8

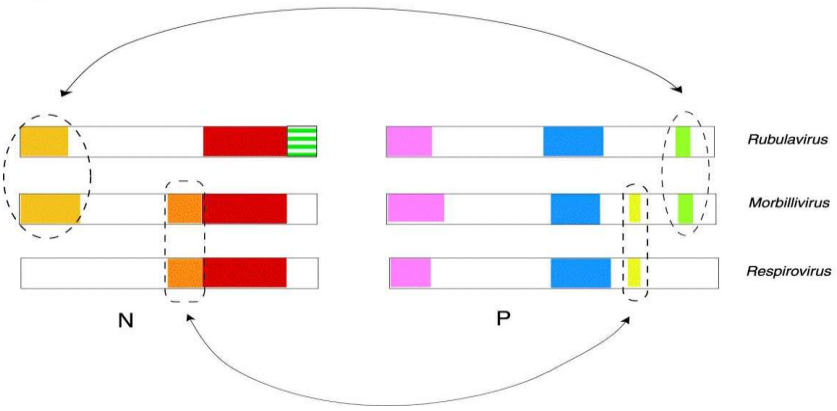

Figure 8. Etude de l'évolution corrélée des régions de la nucléoprotéine et la phosphoprotéine.
Seul le principe est présenté ici. Si une région est conservée dans la nucléoprotéine des *Rubulavirus* et des *Morbillivirus* mais pas des *Respirovirus*, et qu'on observe un motif similaire (**corrélé**) de conservation dans la phosphoprotéine (régions entourées d'un ovale), il est possible que les deux régions interagissent. Bien entendu, cette approche s'applique aussi aux sites d'auto-interaction d'une protéine, N en l'occurrence. Un motif différent de corrélation est présenté pour les *Morbillivirus* et les *Rubulavirus* (régions entourées d'un rectangle). L'étude plus fine de substitutions corrélées au niveau des acides aminés individuels est possible grâce aux logiciels CRASP (Afonnikov *et coll.*, 2001) ou Plotcorr (Pazos *et coll.*, 1997a; Pazos *et coll.*, 1997b).

Le même principe est appliqué à l'ARN (avec beaucoup plus de facilité, étant donné qu'il y a 4 bases au lieu de 20 acides aminés) pour prédire les bases interagissant entre elles au sein d'éléments conservés de structure secondaire, ou d'autres types d'éléments fonctionnels. Il porte alors le nom d'étude de la *covariation* ou des *mutations corrélées* de l'ARN.

À titre d'exemple, la région 295-401 de la nucléoprotéine des *Rubulavirus* (voir figure 4 de la revue) possède un rôle fonctionnel différent de la région correspondante chez les *Respirovirus* et les *Morbillivirus*, puisqu'elle est impliquée dans la liaison de N^{NUC} à P (Nishio *et coll.*, 1999). Un examen d'un alignement multiple de N chez les *Paramyxovirinae* (non montré) indique immédiatement que les aa 295-360 sont hautement conservés, au contraire de la région formée par les aa 361-400. Ainsi, c'est plus probablement cette dernière région qui est impliquée dans l'interaction N^{NUC}-P chez les *Rubulavirus*. Elle est probablement exposée en surface de N chez tous les *Paramyxovirinae*, même si son rôle diffère. Cette analyse est très bien corrélée avec nos études de protéolyse limitée sur N (Karlin *et coll.*, 2002) ainsi qu'avec des études de délétion sur N_{CORE} (Bankamp *et coll.*, 1996; Buchholz *et coll.*, 1993; Liston *et coll.*, 1997). Celles-ci indiquent qu'il existe 3 régions distinctes dans N, même si toutes sont nécessaires à l'auto-assemblage de N : 2 régions dont la délétion provoque la formation d'agrégats non structurés (aa 1-100 environ et aa 385-400) (Bankamp *et coll.*, 1996; Buchholz *et coll.*, 1993; Liston *et coll.*, 1997) - ces régions sont sans doute exposées dans la nucléocapside - et la $3^{ème}$ région (aa 100-385) qui est indispensable à l'auto-association de N. Une idée plus précise de la position de la région 361-400 (qui contient ou est sans doute très proche d'un site de liaison à l'ARN (Myers *et coll.*, 1999)) est proposée en figure 9. La comparaison avec des nucléoprotéines de virus plus éloignés dont seulement quelques caractères comme certains éléments de structure secondaire sont conservés, tels les *Pneumovirinae* ou même les *Rhabdoviridae* et *Filoviridae* (Barr *et coll.*, 1991) apportera un complément utile à cette approche.

Figure 9

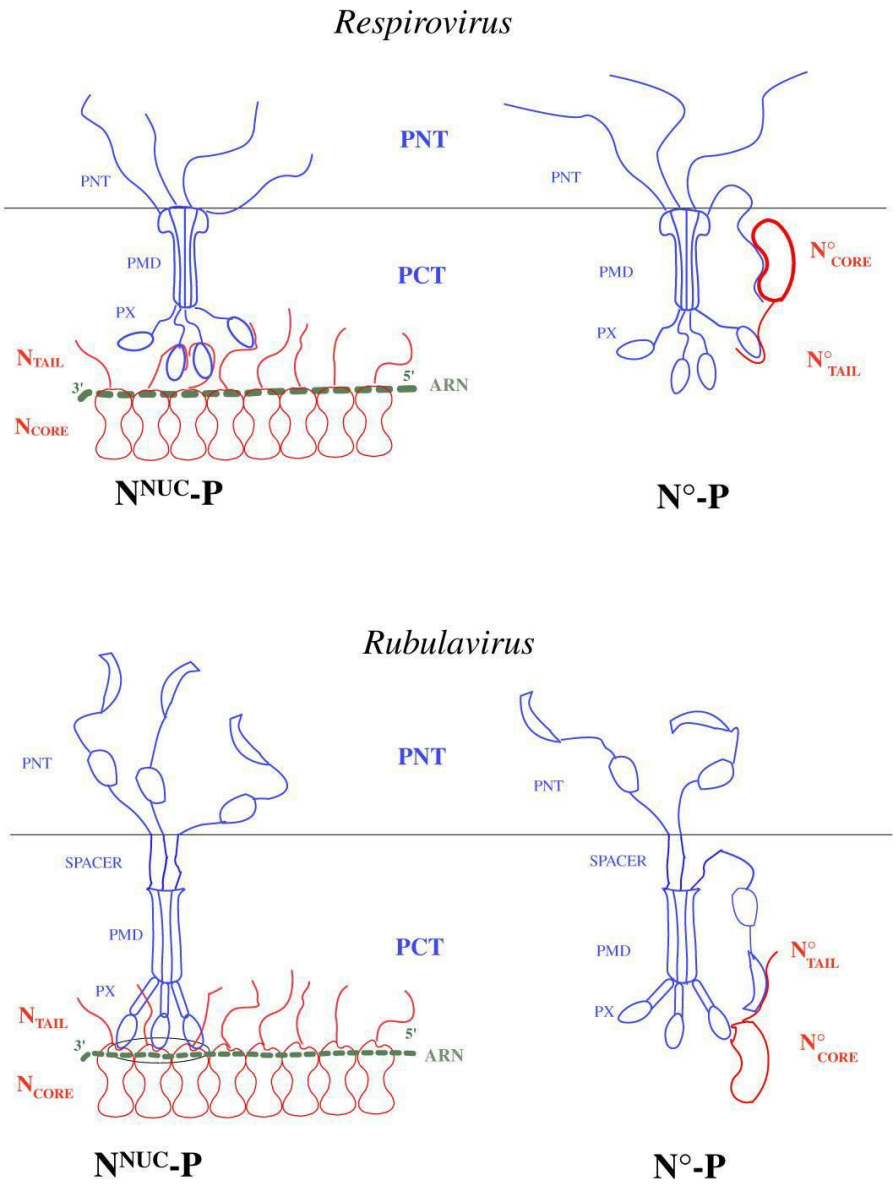

Figure 9. Comparaison des complexes N-P chez les *Rubulaviruses* et les *Respiroviruses*.

Les régions globulaires sont entourées, et les régions non structurées sont figurées par des lignes. Pour les *Rubulavirus*, une partie de PNT est représentée comme étant globulaire, et une partie comme non structurée, en l'absence de données expérimentales (voir revue). La position de l'ARN près du site de liaison de N avec P est figurée par analogie avec les *Rhabdoviridae*.

Le positionnement présumé de la région 361-400 de N des *Rubulavirus* et son interaction avec P sont montrés (ovale noir). Noter la différence avec les *Respirovirus*, pour lesquels l'interaction de P avec la nucléocapside se fait par l'intermédiaire de N_{TAIL}. Voir texte page précédente pour les implications possibles de cette organisation.

Application de l'étude des substitutions corrélées aux régions désordonnées de N et P

L'étude des substitutions corrélées permet d'extraire des informations à partir des régions *variables* des protéines, alors que les biologistes se fondent le plus souvent sur l'étude des régions *conservées*.

À ce titre, cette approche constitue un outil d'étude idéal des régions désordonnées (et très variables) de N et de P, en particulier chez les *Morbillivirus*, pour lesquels on peut établir un alignement multiple très fiable (voir commentaire de la revue). Soulignons une fois encore l'intérêt des virus à ARN pour lesquels il existe de nombreuses séquences homologues, et qui évoluent très rapidement à cause du taux d'erreur élevée des polymérases ARN-dépendantes. Il devrait ainsi être possible d'identifier des régions fonctionnelles dans les régions variables de N et P.

Agrégation de P

Un des principaux obstacles rencontrés au cours de cette thèse est le fait que la phosphoprotéine du virus de la rougeole **s'agrège** quand elle est produite en bactéries, bien qu'elle soit soluble. Pourtant, la fusion du domaine de multimérisation de P à une protéine monomérique provoque la formation de trimères ou de tétramères (Curran *et coll.*, 1995a). Une question lancinante était donc de savoir si P se trouvait agrégée dans les bactéries, ou s'il s'agissait d'un artefact lié à la purification de P, cette dernière s'agrégeant suite à une dénaturation mineure (on peut aussi penser à l'influence de l'épitope hexahistidine, ou encore de cystéines qui établiraient des ponts disulfures, solutions apparemment exclues par l'expérimentation).

Cette question revêt une importance particulière car l'étude des protéines agrégées, très difficile dans l'état actuel des techniques, est d'une grande importance en médecine (à l'exemple de la protéine prion, ou de l'α-synucléine impliquée dans la maladie de Parkinson). En particulier, le groupe de A. Fink à Santa Cruz (CA, USA) qui comprend notamment Vladimir Uversky (Uversky, 2002), est en train de montrer toute l'importance des états non structurés des protéines dans leur agrégation (Uversky *et coll.*, 2001). En première instance, il était donc tentant d'imaginer que l'agrégation de P était causée par la première moitié de P, PNT, qui est intrinsèquement désordonnée. Cette hypothèse est improbable puisque l'autre moitié de P, PCT, s'agrège aussi. En s'appuyant sur notre étude de l'organisation modulaire de P (voir la revue), il devrait être possible de purifier les différents domaines de PCT et d'établir lequel cause l'agrégation de P. Si tous les domaines sont agrégés, il s'agit très probablement d'un artefact trivial (voir ci-dessus).

Nous avons tout d'abord pensé que P pourrait former des agrégats amyloïdes ; pourtant, le spectre de P (figure 10) indique la présence de régions non structurées (sans doute PNT ?)

mais aussi probablement de régions de structure secondaire α-hélicoïdale (conformément à la prédiction de structure secondaire) - alors que les protéines amyloïdes s'agrègent via des feuillets β. Ceci suggère que P aurait une conformation similaire à celle attendue, et que son agrégation pourrait provenir d'un problème mineur lors de la purification.

Figure 10

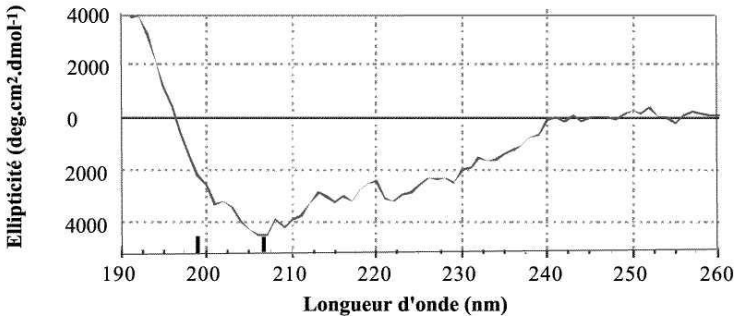

Figure 10. Spectre de P en dichroïsme circulaire.

Le spectre a été établi à une concentration de 0.3 mg/ml, en NaP 10 mM pH 7. L'ellipticité négative à 198 nm ainsi que le minimum à 208 nm (marques noires) indiquent respectivement la présence de régions non structurées, et de structure secondaire α-hélicoïdale.

Enfin, le fait qu'au sein de lysats bactériens P puisse se lier à N (voir l'article de la présente thèse), avec une stœchiométrie précise (P_1 N_1) en conditions d'excès de P (D. Karlin et K. Johansson, résultats non publiés) incite fortement à penser que P est bien replié avant son extraction des bactéries. Le mystère concernant l'agrégation de P lors de sa purification reste entier, et il serait très intéressant de poursuivre cette piste de recherche. S'il ne s'agit pas d'une explication triviale, les grands spécialistes de l'agrégation et des protéines désordonnées que sont V. Uversky et A. Fink. seront sans doute intéressés par ce problème et pourront s'y attaquer avec des moyens adéquats (par exemple en déterminant si P forme des fibres).

P complexé à N° peut-il se lier à N^{NUC} ?

Chez SeV, P complexé à N° (avec une stœchiométrie $N_1°P_4$, J. Curran, communication personnelle), possède 3 régions C-terminales libres de se lier à d'autres molécules de N^{NUC} (les domaines XD). Il est légitime de se demander si celles-ci ne se lient pas à la nucléocapside matrice lors de la réplication. Dès que nous disposerons des deux complexes N°-P et N^{NUC}-P, il devrait être possible d'étudier une telle question, grâce à la combinaison d'anticorps monoclonaux dirigés contre différents épitopes naturels ou artificiels de N et P (par exemple, N peut être produit avec un épitope Flag et P avec un épitope hexahistidine).

Une autre problématique est : comment P se dissocie-t-il de N° pour permettre l'encapsidation ? La figure 11 suppose qu'au sein du complexe N°-P le domaine XD complexé à la queue C-terminale de N° pourrait (après une modification de cette dernière, induite par L ?) relâcher la queue de N° et se lier à la nucléocapside *matrice*. Cette question nécessite de reconstituer un complexe de réplication, un but encore lointain.

Figure 11

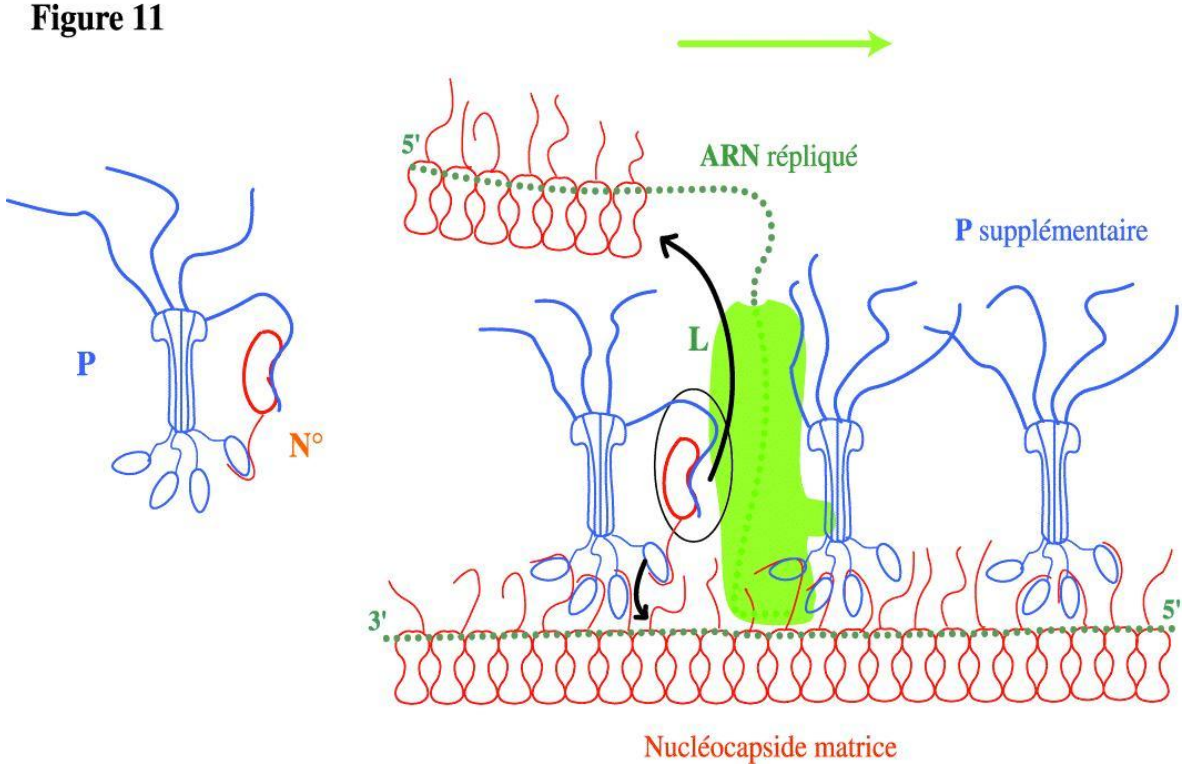

Figure 11. Modèle d'interaction tripartite de N, P, et de la polymérase durant la réplication.
Polymérase d'un *Respirovirus* en cours de réplication d'un ARN génomique. Les conventions utilisées sont les mêmes qu'en Figure 9. La flèche verte indique le sens de progression de la polymérase. L (grande forme verte) est liée à un tétramère de P à l'extrémité C-terminale du domaine de multimerisation de celui-ci (PMD). Une molécule supplémentaire de P, non liée à L, est représentée à droite. On peut voir l'ARN néosynthétisé émerger de la polymérase. Il est déjà partiellement encapsidé, bien qu'il n'y ait pas de preuve que l'encapsidation commence à l'extrémité 5', comme nous l'avons représenté. Le substrat de la polymérase pour l'encapsidation, N°-P, est mis en contact avec le complexe (N:ARN) nouvellement répliqué par L via un contact PNT-L, conduisant à la formation un complexe tripartite entre N°, P, et la polymérase (entouré d'un cercle). P au sein de N°-P est représenté lié à la nucléocapside *matrice* via ses "bras" libres (PX), bien qu'il n'y ait pas de preuve expérimentale en faveur de cette hypothèse. Cependant, P au sein de N°-P ne se lie sans doute pas au complexe (N:ARN) *nouvellement répliqué*, car la partie N-terminale de N suffit à l'encapsidation (Curran *et coll.*, 1993). Il est tentant d'imaginer que la proximité de la nucléocapside puisse provoquer le relargage de N° par PX. Ceci provoquerait l'incorporation de N° à la nucléocapside en cours d'assemblage (longue flèche) ainsi que la liaison des PX libres à la matrice (N:ARN) (flèche courte).

Conceptions d'agents anti-viraux

Les études structurales sur un virus qui appartient à l'ordre des *Mononegavirales* ont un grand impact, à cause de la similitude des virus au sein de cette "super-famille" qui contient plus d'une vingtaine de virus pathogènes très répandus, redoutables, ou les deux. Ainsi, mes résultats sur la flexibilité structurale des phosphoprotéines des *Paramyxoviridae* peuvent manifestement être étendus aux *Rhabdoviridae*.

Un des buts initiaux de ma thèse était de résoudre la structure tridimensionnelle de N, P, et L, afin de permettre la conception rationnelle d'antiviraux. Je pense que ma thèse aura montré que la phosphoprotéine ne saurait constituer une cible pour les techniques *classiques* de conception rationnelle d'agents inhibiteurs de la réplication virale, fondées sur des connaissances structurales de protéines cibles. Cela n'exclut bien sûr pas un avenir où les régions désordonnées des protéines (probablement présentes dans 40% des protéines humaines (Dunker and Obradovic, 2001)) constitueraient des cibles thérapeutiques pour des antiviraux dont les modalités d'action et la conception restent encore à définir …

CONCLUSION

Dans le cours de ma thèse, j'ai étudié la phosphoprotéine et la nucléoprotéine du virus de la rougeole, qui jouent un rôle crucial au sein du complexe de réplication viral. J'ai étudié les déterminants de la polymérisation de N et identifié un variant de N qui présente un intérêt pour la cristallographie. Mon travail a aussi permis la mise au point d'un système expérimental d'étude des complexes formés par N et P. Par ailleurs, j'ai démontré que le domaine N-terminal de la phosphoprotéine du virus de la rougeole est intrinsèquement désordonné, puis étendu ce résultat à la famille de ce virus, les *Paramyxoviridae*, ainsi qu'à une famille voisine, les *Rhabdoviridae*. J'ai ensuite montré l'importance du désordre structural au sein du complexe réplicatif des *Paramyxoviridae* (à la fois par son abondance et d'un point de vue fonctionnel). Au-delà de leur intérêt pratique immédiat, ces résultats indiquent que la machinerie réplicative de ces virus présente un intérêt comme système modèle de l'étude du désordre structural dans les protéines. De plus, ces résultats soulèvent de nombreuses questions et sont donc un prélude à une refonte de notre vision du complexe réplicatif de ces virus dont l'étude revêt une importance majeure en santé humaine.

BIBLIOGRAPHIE

Afonnikov, D. A., Oshchepkov, D. Y., and Kolchanov, N. A. (2001). Detection of conserved physico-chemical characteristics of proteins by analyzing clusters of positions with co-ordinated substitutions. *Bioinformatics* **17**(11)**,** 1035-46.

Baker, S. C., and Moyer, S. A. (1988). Encapsidation of Sendai virus genome RNAs by purified NP protein during in vitro replication. *J Virol* **62**(3)**,** 834-8.

Bankamp, B., Horikami, S. M., Thompson, P. D., Huber, M., Billeter, M., and Moyer, S. A. (1996). Domains of the measles virus N protein required for binding to P protein and self-assembly. *Virology* **216**(1)**,** 272-7.

Barr, J., Chambers, P., Pringle, C. R., and Easton, A. J. (1991). Sequence of the major nucleocapsid protein gene of pneumonia virus of mice: sequence comparisons suggest structural homology between nucleocapsid proteins of pneumoviruses, paramyxoviruses, rhabdoviruses and filoviruses. *J Gen Virol* **72**(Pt 3)**,** 677-85.

Baudin, F., Bach, C., Cusack, S., and Ruigrok, R. W. (1994). Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *Embo J* **13**(13)**,** 3158-65.

Bohn, W., Ciampor, F., Rutter, R., and Mannweiler, K. (1990). Localization of nucleocapsid associated polypeptides in measles virus- infected cells by immunogold labelling after resin embedding. *Arch Virol* **114**(1-2)**,** 53-64.

Bousse, T., Takimoto, T., Matrosovich, T., and Portner, A. (2001). Two regions of the P protein are required to be active with the L protein for human parainfluenza virus type 1 RNA polymerase activity. *Virology* **283**(2)**,** 306-14.

Bowman, M. C., Smallwood, S., and Moyer, S. A. (1999). Dissection of individual functions of the Sendai virus phosphoprotein in transcription. *J. Virol.* **73**(8)**,** 6474-83.

Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T., Oldfield, C. J., Williams, C. J., and Dunker, A. K. (2002). Titre non communiqué. *J. Mol. Evol.* ***in press.***

Buchholz, C. J., Spehner, D., Drillien, R., Neubert, W. J., and Homann, H. E. (1993). The conserved N-terminal region of Sendai virus nucleocapsid protein NP is required for nucleocapsid assembly. *J Virol* **67**(10)**,** 5803-12.

Calain, P., and Roux, L. (1993). The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. *J Virol* **67**(8)**,** 4822-30.

Castaneda, S. J., and Wong, T. C. (1990). Leader sequence distinguishes between translatable and encapsidated measles virus RNAs. *J Virol* **64**(1)**,** 222-30.

Chandrika, R., Myers, T., and Moyer, S. A. (1995). Measles virus nucleocapsid protein can function in Sendai virus defective interfering particle genome synthesis in vitro. *Virology* **206**(1)**,** 777-82.

Chastel, C. (1992). Le virus sigma : un virus d'insecte au pouvoir pathogène étrange. *Histoire des virus - de la variole au sida- Editions Boubée, Paris.***,** 173-85.

Chinchar, V. G., and Portner, A. (1981). Functions of Sendai virus nucleocapsid polypeptides: enzymatic activities in nucleocapsids following cleavage of polypeptide P by Staphylococcus aureus protease V8. *Virology* **109**(1)**,** 59-71.

Choudhary, S., De, B. P., and Banerjee, A. K. (2000). Specific phosphorylated forms of glyceraldehyde 3-phosphate dehydrogenase associate with human parainfluenza virus type 3 and inhibit viral transcription in vitro. *J Virol* **74**(8)**,** 3634-41.

Coronel, E. C., Takimoto, T., Murti, K. G., Varich, N., and Portner, A. (2001). Nucleocapsid incorporation into parainfluenza virus is regulated by specific interaction with matrix protein. *J Virol* **75**(3)**,** 1117-23.

Cortese, C. K., Feller, J. A., and Moyer, S. A. (2000). Mutations in domain V of the Sendai virus L polymerase protein uncouple transcription and replication and differentially affect replication in vitro and in vivo. *Virology* **277**(2)**,** 387-96.

Curran, J. (1996). Reexamination of the Sendai virus P protein domains required for RNA synthesis: a possible supplemental role for the P protein. *Virology* **221**(1)**,** 130-40.

Curran, J. (1998). A role for the Sendai virus P protein trimer in RNA synthesis. *J. Virol.* **72**(5)**,** 4274-80.

Curran, J., Boeck, R., Lin-Marq, N., Lupas, A., and Kolakofsky, D. (1995a). Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils. *Virology* **214**(1)**,** 139-49.

Curran, J., Homann, H., Buchholz, C., Rochat, S., Neubert, W., and Kolakofsky, D. (1993). The hypervariable C-terminal tail of the Sendai paramyxovirus nucleocapsid protein is required for template function but not for RNA encapsidation. *J. Virol.* **67**(7)**,** 4358-64.

Curran, J., and Kolakofsky, D. (1999). Replication of paramyxoviruses. *Adv. Virus Res.* **54,** 403-22.

Curran, J., Marq, J. B., and Kolakofsky, D. (1995b). An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J. Virol.* **69**(2)**,** 849-55.

Curran, J., Pelet, T., and Kolakofsky, D. (1994). An acidic activation-like domain of the Sendai virus P protein is required for RNA synthesis and encapsidation. *Virology* **202**(2)**,** 875-84.

Curran, J. A., and Kolakofsky, D. (1991). Rescue of a Sendai virus DI genome by other parainfluenza viruses: implications for genome replication. *Virology* **182**(1)**,** 168-76.

Das, T., and Banerjee, A. K. (1993). Expression of the vesicular stomatitis virus nucleocapsid protein gene in Escherichia coli: analysis of its biological activity in vitro. *Virology* **193**(1)**,** 340-7.

Das, T., Mathur, M., Gupta, A. K., Janssen, G. M., and Banerjee, A. K. (1998). RNA polymerase of vesicular stomatitis virus specifically associates with translation elongation factor-1 alphabetagamma for its activity. *Proc Natl Acad Sci U S A* **95**(4)**,** 1449-54.

De, B. P., and Banerjee, A. K. (1999). Involvement of actin microfilaments in the transcription/replication of human parainfluenza virus type 3: possible role of actin in other viruses. *Microsc Res Tech* **47**(2)**,** 114-23.

Deshpande, K. L., and Portner, A. (1984). Structural and functional analysis of Sendai virus nucleocapsid protein NP with monoclonal antibodies. *Virology* **139**(1)**,** 32-42.

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001). Intrinsically disordered protein. *J. Mol. Graph. Model.* **19**(1)**,** 26-59.

Dunker, A. K., and Obradovic, Z. (2001). The protein trinity--linking function and disorder. *Nat Biotechnol* **19**(9)**,** 805-6.

Egelman, E. H., Wu, S. S., Amrein, M., Portner, A., and Murti, G. (1989). The Sendai virus nucleocapsid exists in at least four different helical states. *J Virol* **63**(5)**,** 2233-43.

Einberger, H., Mertz, R., Hofschneider, P. H., and Neubert, W. J. (1990). Purification, renaturation, and reconstituted protein kinase activity of the Sendai virus large (L) protein: L protein phosphorylates the NP and P proteins in vitro. *J Virol* **64**(9)**,** 4274-80.

Erlenhoefer, C., Wurzer, W. J., Loffler, S., Schneider-Schaulies, S., ter Meulen, V., and Schneider-Schaulies, J. (2001). CD150 (SLAM) is a receptor for measles virus but is not involved in viral contact-mediated proliferation inhibition. *J Virol* **75**(10)**,** 4499-505.

Escoffier, C., Manie, S., Vincent, S., Muller, C. P., Billeter, M., and Gerlier, D. (1999). Nonstructural C protein is required for efficient measles virus replication in human peripheral blood cells. *J Virol* **73**(2)**,** 1695-8.

Feller, J. A., Smallwood, S., Horikami, S. M., and Moyer, S. A. (2000). Mutations in conserved domains IV and VI of the large (L) subunit of the sendai virus RNA polymerase give a spectrum of defective RNA synthesis phenotypes. *Virology* **269**(2)**,** 426-39.

Ferron, F., Longhi, S., Henrissat, B., and Canard, B. (2002). Viral RNA-polymerases - a predicted 2-'O ribose methyltransferase domain shared by all *Mononegavirales*. *TIBS* **27**(5)**,** 222-4.

Galinski, M. S. (1991). Paramyxoviridae: transcription and replication. *Adv. Virus Res.* **39,** 129-62.

Gill, D. S., Takai, S., Portner, A., and Kingsbury, D. W. (1988). Mapping of antigenic domains of Sendai virus nucleocapsid protein expressed in Escherichia coli. *J Virol* **62**(12)**,** 4805-8.

Gombart, A. F., Hirano, A., and Wong, T. C. (1992). Expression and properties of the V protein in acute measles virus and subacute sclerosing panencephalitis virus strains. *Virus Res* **25**(1-2)**,** 63-78.

Gombart, A. F., Hirano, A., and Wong, T. C. (1993). Conformational maturation of measles virus nucleocapsid protein. *J. Virol.* **67**(7)**,** 4133-41.

Green, T. J., Macpherson, S., Qiu, S., Lebowitz, J., Wertz, G. W., and Luo, M. (2000). Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. *J Virol* **74**(20)**,** 9515-24.

Griffin, D. E. (2001). Measles Virus. *In "Fields Virology" (B.N. Fields, D.M. Knipe, and P.M. Howley, Eds.), 4th ed., pp 1401-1441 Lippincott-Raven, Philadelphia, PA.*

Grogan, C. C., and Moyer, S. A. (2001). Sendai virus wild-type and mutant C proteins show a direct correlation between L polymerase binding and inhibition of viral RNA synthesis. *Virology* **288**(1)**,** 96-108.

Grossmann, J. G., Sharff, A. J., O'Hare, P., and Luisi, B. (2001). Molecular shapes of transcription factors TFIIB and VP16 in solution: implications for recognition. *Biochemistry* **40**(21)**,** 6267-74.

Gubbay, O., Curran, J., and Kolakofsky, D. (2001). Sendai virus genome synthesis and assembly are coupled: a possible mechanism to promote viral RNA polymerase processivity. *J. Gen. Virol.* **82**(Pt 12)**,** 2895-903.

Heggeness, M. H., Scheid, A., and Choppin, P. W. (1980). Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration. *Proc Natl Acad Sci U S A* **77**(5)**,** 2631-5.

Heggeness, M. H., Scheid, A., and Choppin, P. W. (1981). The relationship of conformational changes in the Sendai virus nucleocapsid to proteolytic cleavage of the NP polypeptide. *Virology* **114**(2)**,** 555-62.

Hirano, A., Ayata, M., Wang, A. H., and Wong, T. C. (1993). Functional analysis of matrix proteins expressed from cloned genes of measles virus variants that cause subacute sclerosing panencephalitis reveals a common defect in nucleocapsid binding. *J Virol* **67**(4)**,** 1848-53.

Hirano, A., Wang, A. H., Gombart, A. F., and Wong, T. C. (1992). The matrix proteins of neurovirulent subacute sclerosing panencephalitis virus and its acute measles virus progenitor are functionally different. *Proc Natl Acad Sci U S A* **89**(18)**,** 8745-9.

Holmes, D. E., and Moyer, S. A. (2002). The phosphoprotein (p) binding site resides in the N terminus of the L polymerase subunit of sendai virus. *J Virol* **76**(6)**,** 3078-83.

Horikami, S. M., and Moyer, S. A. (1995). Structure, transcription, and replication of measles virus. *Curr Top Microbiol Immunol* **191,** 35-50.

Horikami, S. M., Smallwood, S., Bankamp, B., and Moyer, S. A. (1994). An amino-proximal domain of the L protein binds to the P protein in the measles virus RNA polymerase complex. *Virology* **205**(2)**,** 540-5.

Hsu, E. C., Iorio, C., Sarangi, F., Khine, A. A., and Richardson, C. D. (2001). CDw150(SLAM) is a receptor for a lymphotropic strain of measles virus and may account for the immunosuppressive properties of this virus. *Virology* **279**(1)**,** 9-21.

Jonson, P. H., and Petersen, S. B. (2001). A critical view on conservative mutations. *Protein Eng* **14**(6)**,** 397-402.

Karlen, A. (1995). "Plague's progress. A social history of man and disease", Indigo, London, UK.

Karlin, D., Longhi, S., and Canard, B. (2002). Substitution of two residues in the measles virus nucleoprotein results in an impaired self-association **Submitted for publication**.

Kolakofsky, D., Pelet, T., Garcin, D., Hausmann, S., Curran, J., and Roux, L. (1998). Paramyxovirus RNA synthesis and the requirement for hexamer genome length: the rule of six revisited. *J Virol* **72**(2)**,** 891-9.

Lamb, R. A., and Choppin, P. W. (1977a). The synthesis of Sendai virus polypeptides in infected cells. II. Intracellular distribution of polypeptides. *Virology* **81**(2)**,** 371-81.

Lamb, R. A., and Choppin, P. W. (1977b). The synthesis of Sendai virus polypeptides in infected cells. III. Phosphorylation of polypeptides. *Virology* **81**(2)**,** 382-97.

Lamb, R. A., and Kolakofsky, D. (2001). *Paramyxoviridae* : The Viruses and Their Replication. *In "Fields Virology" (B.N. Fields, D.M. Knipe, and P.M. Howley, Eds.), 4th ed., pp 1305-1340. Lippincott-Raven, Philadelphia, PA.*

Latorre, P., Kolakofsky, D., and Curran, J. (1998). Sendai virus Y proteins are initiated by a ribosomal shunt. *Mol Cell Biol* **18**(9)**,** 5021-31.

Lecellier, C. H., Petit, C., and Saïb, A. (2002). Le cytosquelette et les virus, ou comment se déplacer dans une cellule. *Virologie* **6**(1)**,** 9-18.

Leopardi, R., Hukkanen, V., Vainionpaa, R., and Salmi, A. A. (1993). Cell proteins bind to sites within the 3' noncoding region and the positive-strand leader sequence of measles virus RNA. *J. Virol.* **67**(2)**,** 785-90.

Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999). Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform. Ser. Workshop Genome Inform.* **10,** 30-40.

Liston, P., Batal, R., DiFlumeri, C., and Briedis, D. J. (1997). Protein interaction domains of the measles virus nucleocapsid protein (NP). *Arch Virol* **142**(2)**,** 305-21.

Liston, P., and Briedis, D. J. (1995). Ribosomal frameshifting during translation of measles virus P protein mRNA is capable of directing synthesis of a unique protein. *J. Virol.* **69**(11)**,** 6742-50.

Lund, G. A., Tyrrell, D. L., Bradley, R. D., and Scraba, D. G. (1984). The molecular length of measles virus RNA and the structural organization of measles nucleocapsids. *J Gen Virol* **65**(Pt 9)**,** 1535-42.

Manie, S. N., Debreyne, S., Vincent, S., and Gerlier, D. (2000). Measles virus structural components are enriched into lipid raft microdomains: a potential cellular location for virus assembly. *J Virol* **74**(1)**,** 305-11.

Marriott, A. C., and Easton, A. J. (1999). Reverse genetics of the Paramyxoviridae. *Adv. Virus Res.* **53,** 321-40.

Miroux, B., and Walker, J. E. (1996). Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**(3)**,** 289-98.

Mottet, G., Muller, V., and Roux, L. (1999). Characterization of Sendai virus M protein mutants that can partially interfere with virus particle production. *J Gen Virol* **80**(Pt 11)**,** 2977-86.

Moyer, S. A., Baker, S. C., and Horikami, S. M. (1990). Host cell proteins required for measles virus reproduction. *J. Gen. Virol.* **71**(Pt 4)**,** 775-83.

Murphy, S. K., Ito, Y., and Parks, G. D. (1998). A functional antigenomic promoter for the paramyxovirus simian virus 5 requires proper spacing between an essential internal segment and the 3' terminus. *J Virol* **72**(1)**,** 10-9.

Murphy, S. K., and Parks, G. D. (1999). RNA replication for the paramyxovirus simian virus 5 requires an internal repeated (CGNNNN) sequence motif. *J Virol* **73**(1)**,** 805-9.

Myers, T. M., and Moyer, S. A. (1997). An amino-terminal domain of the Sendai virus nucleocapsid protein is required for template function in viral RNA synthesis. *J. Virol.* **71**(2)**,** 918-24.

Myers, T. M., Smallwood, S., and Moyer, S. A. (1999). Identification of nucleocapsid protein residues required for Sendai virus nucleocapsid formation and genome replication. *J. Gen. Virol.* **80**(Pt 6)**,** 1383-91.

Nagai, Y. (1999). Paramyxovirus replication and pathogenesis. Reverse genetics transforms understanding. *Rev Med Virol* **9**(2)**,** 83-99.

Naniche, D., Varior-Krishnan, G., Cervoni, F., Wild, T. F., Rossi, B., Rabourdin-Combe, C., and Gerlier, D. (1993). Human membrane cofactor protein (CD46) acts as a cellular receptor for measles virus. *J Virol* **67**(10)**,** 6025-32.

Nishio, M., Tsurudome, M., Ito, M., and Ito, Y. (2000). Mapping of domains on the human parainfluenza type 2 virus P and NP proteins that are involved in the interaction with the L protein. *Virology* **273**(2)**,** 241-7.

Nishio, M., Tsurudome, M., Ito, M., Kawano, M., Kusagawa, S., Komada, H., and Ito, Y. (1999). Mapping of domains on the human parainfluenza virus type 2 nucleocapsid protein (NP) required for NP-phosphoprotein or NP-NP interaction. *J Gen Virol* **80**(Pt 8)**,** 2017-22.

Nishio, M., Tsurudome, M., Ito, M., Watanabe, N., Kawano, M., Komada, H., and Ito, Y. (1997). Human parainfluenza virus type 2 phosphoprotein: mapping of monoclonal antibody epitopes and location of the multimerization domain. *J. Gen. Virol.* **78**(Pt 6)**,** 1303-8.

Nishio, M., Tsurudome, M., Kawano, M., Watanabe, N., Ohgimoto, S., Ito, M., Komada, H., and Ito, Y. (1996). Interaction between nucleocapsid protein (NP) and phosphoprotein (P) of human parainfluenza virus type 2: one of the two NP binding sites on P is essential for granule formation. *J. Gen. Virol.* **77**(Pt 10)**,** 2457-63.

O'Reilly, E. K., and Kao, C. C. (1998). Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. *Virology* **252**(2)**,** 287-303.

Oglesbee, M., Ringler, S., and Krakowka, S. (1990). Interaction of canine distemper virus nucleocapsid variants with 70K heat-shock proteins. *J Gen Virol* **71**(Pt 7)**,** 1585-90.

Oglesbee, M. J., Liu, Z., Kenney, H., and Brooks, C. L. (1996). The highly inducible member of the 70 kDa family of heat shock proteins increases canine distemper virus polymerase activity. *J. Gen. Virol.* **77**(Pt 9)**,** 2125-35.

Omer, M. I. (1999). Measles: a disease that has to be eradicated. *Ann Trop Paediatr* **19**(2)**,** 125-34.

Ono, N., Tatsuo, H., Hidaka, Y., Aoki, T., Minagawa, H., and Yanagi, Y. (2001). Measles viruses on throat swabs from measles patients use signaling lymphocytic activation molecule (CDw150) but not CD46 as a cellular receptor. *J Virol* **75**(9)**,** 4399-401.

Parks, C. L., Lerch, R. A., Walpita, P., Sidhu, M. S., and Udem, S. A. (1999). Enhanced measles virus cDNA rescue and gene expression after heat shock. *J Virol* **73**(5)**,** 3560-6.

Parks, C. L., Lerch, R. A., Walpita, P., Wang, H. P., Sidhu, M. S., and Udem, S. A. (2001). Comparison of predicted amino acid sequences of measles virus strains in the Edmonston vaccine lineage. *J Virol* **75**(2)**,** 910-20.

Patterson, J. B., Thomas, D., Lewicki, H., Billeter, M. A., and Oldstone, M. B. (2000). V and C proteins of measles virus function as virulence factors in vivo. *Virology* **267**(1)**,** 80-9.

Pazos, F., Helmer-Citterich, M., Ausiello, G., and Valencia, A. (1997a). Correlated mutations contain information about protein-protein interaction. *J Mol Biol* **271**(4)**,** 511-23.

Pazos, F., Olmea, O., and Valencia, A. (1997b). A graphical interface for correlated mutations and other protein structure prediction methods. *Comput Appl Biosci* **13**(3)**,** 319-21.

Poch, O., Blumberg, B. M., Bougueleret, L., and Tordo, N. (1990). Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *J Gen Virol* **71**(Pt 5)**,** 1153-62.

Raj, P. A., and Dentino, A. R. (2002). Current status of defensins and their role in innate and adaptive immunity. *FEMS Microbiol Lett* **206**(1)**,** 9-18.

Ray, J., and Fujinami, R. S. (1987). Characterization of in vitro transcription and transcriptional products of measles virus. *J Virol* **61**(11)**,** 3381-7.

Reutter, G. L., Cortese-Grogan, C., Wilson, J., and Moyer, S. A. (2001). Mutations in the measles virus C protein that up regulate viral RNA synthesis. *Virology* **285**(1)**,** 100-9.

Ruff, T. A. (1999). Immunisation strategies for viral diseases in developing countries. *Rev Med Virol* **9**(2)**,** 121-38.

Schoehn, G., Iseni, F., Mavrakis, M., Blondel, D., and Ruigrok, R. W. (2001). Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. *J. Virol.* **75**(1)**,** 490-8.

Smallwood, S., Easson, C. D., Feller, J. A., Horikami, S. M., and Moyer, S. A. (1999). Mutations in conserved domain II of the large (L) subunit of the Sendai virus RNA polymerase abolish RNA synthesis. *Virology* **262**(2)**,** 375-83.

Spehner, D., Drillien, R., and Howley, P. M. (1997). The assembly of the measles virus nucleoprotein into nucleocapsid-like particles is modulated by the phosphoprotein. *Virology* **232**(2)**,** 260-8.

Stillman, E. A., and Whitt, M. A. (1999). Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription. *J Virol* **73**(9)**,** 7199-209.

Stricker, R., Mottet, G., and Roux, L. (1994). The Sendai virus matrix protein appears to be recruited in the cytoplasm by the viral nucleocapsid to function in viral assembly and budding. *J Gen Virol* **75**(Pt 5)**,** 1031-42.

Suryanarayana, K., Baczko, K., ter Meulen, V., and Wagner, R. R. (1994). Transcription inhibition and other properties of matrix proteins expressed by M genes cloned from measles viruses and diseased human brain tissue. *J Virol* **68**(3)**,** 1532-43.

Takeda, M., Kato, A., Kobune, F., Sakata, H., Li, Y., Shioda, T., Sakai, Y., Asakawa, M., and Nagai, Y. (1998). Measles virus attenuation associated with transcriptional impediment and a few amino acid changes in the polymerase and accessory proteins. *J Virol* **72**(11)**,** 8690-6.

Takimoto, T., Murti, K. G., Bousse, T., Scroggs, R. A., and Portner, A. (2001). Role of matrix and fusion proteins in budding of Sendai virus. *J Virol* **75**(23)**,** 11384-91.

Tanaka, K., Minagawa, H., Xie, M. F., and Yanagi, Y. (2002). The measles virus hemagglutinin downregulates the cellular receptor SLAM (CD150). *Arch Virol* **147**(1)**,** 195-203.

Tapparel, C., Maurice, D., and Roux, L. (1998). The activity of Sendai virus genomic and antigenomic promoters requires a second element past the leader template regions: a motif (GNNNNN)3 is essential for replication. *J Virol* **72**(4)**,** 3117-28.

Tarbouriech, N., Curran, J., Ebel, C., Ruigrok, R. W., and Burmeister, W. P. (2000). On the domain structure and the polymerization state of the sendai virus P protein. *Virology* **266**(1)**,** 99-109.

Tatsuo, H., Ono, N., and Yanagi, Y. (2001). Morbilliviruses use signaling lymphocyte activation molecules (CD150) as cellular receptors. *J Virol* **75**(13)**,** 5842-50.

Tsai, C. J., Ma, B., Sham, Y. Y., Kumar, S., and Nussinov, R. (2001). Structured disorder and conformational selection. *Proteins* **44**(4)**,** 418-27.

Tuckis, J., Smallwood, S., Feller, J. A., and Moyer, S. A. (2002). The C-terminal 88 amino acids of the Sendai virus P protein have multiple functions separable by mutation. *J Virol* **76**(1)**,** 68-77.

Uversky, V. N. (2002). What does it mean to be natively unfolded? *Eur J Biochem* **269**(1)**,** 2-12.

Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000a). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41**(3)**,** 415-427.

Uversky, V. N., Gillespie, J. R., Millett, I. S., Khodyakova, A. V., Vasilenko, R. N., Vasiliev, A. M., Rodionov, I. L., Kozlovskaya, G. D., Dolgikh, D. A., Fink, A. L., Doniach, S., Permyakov, E. A., and Abramov, V. M. (2000b). Zn(2+)-mediated structure formation and compaction of the "natively unfolded" human prothymosin alpha. *Biochem Biophys Res Commun* **267**(2)**,** 663-8.

Uversky, V. N., Gillespie, J. R., Millett, I. S., Khodyakova, A. V., Vasiliev, A. M., Chernovskaya, T. V., Vasilenko, R. N., Kozlovskaya, G. D., Dolgikh, D. A., Fink, A. L., Doniach, S., and Abramov, V. M. (1999). Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH. *Biochemistry* **38**(45)**,** 15009-16.

Uversky, V. N., Li, J., and Fink, A. L. (2001). Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J Biol Chem* **276**(14)**,** 10737-44.

Valsamakis, A., Schneider, H., Auwaerter, P. G., Kaneshima, H., Billeter, M. A., and Griffin, D. E. (1998). Recombinant measles viruses with mutations in the C, V, or F gene have altered growth phenotypes in vivo. *J. Virol.* **72**(10)**,** 7754-61.

Vincent, S., Gerlier, D., and Manie, S. N. (2000). Measles virus assembly within membrane rafts. *J Virol* **74**(21)**,** 9911-5.

Vulliemoz, D., and Roux, L. (2001). "Rule of six": how does the Sendai virus RNA polymerase keep count? *J Virol* **75**(10)**,** 4506-18.

Warnes, A., Fooks, A. R., Dowsett, A. B., Wilkinson, G. W., and Stephenson, J. R. (1995). Expression of the measles virus nucleoprotein gene in Escherichia coli and assembly of nucleocapsid-like structures. *Gene* **160**(2)**,** 173-8.

Watanabe, N., Kawano, M., Tsurudome, M., Nishio, M., Ito, M., Ohgimoto, S., Suga, S., Komada, H., and Ito, Y. (1996). Binding of the V proteins to the nucleocapsid proteins of human parainfluenza type 2 virus. *Med. Microbiol. Immunol. (Berl)* **185**(2)**,** 89-94.

Williams, R. M., Obradovi, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., Takayama, S., Brown, C. J., and Dunker, A. K. (2001). The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac. Symp. Biocomput.***,** 89-100.

Wright, P. E., and Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**(2)**,** 321-31.

Young, J. A. (2001). Virus entry and uncoating. *In "Fields Virology" (B.N. Fields, D.M. Knipe, and P.M. Howley, Eds.), 4th ed., pp 87-103. Lippincott-Raven, Philadelphia, PA.*

Zetina, C. R. (2001). A conserved helix-unfolding motif in the naturally unfolded proteins. *Proteins* **44**(4)**,** 479-83.

REMERCIEMENTS

Une thèse est un travail d'équipe. Je tiens à remercier :

Christian Cambillau pour m'avoir accueilli dans son laboratoire et pour avoir soutenu ma candidature à une bourse de docteur-ingénieur.

Bruno Canard pour m'avoir accueilli dans son équipe, pour sa grande patience, sa compréhension, son exemple, son enthousiasme communicatif, pour m'avoir guidé aux bons moments et pour m'avoir fait confiance aux (autres) bons moments.

Sonia Longhi pour son immense patience, sa très grande gentillesse, son immense disponibilité, son amour des étudiants, son aide constante, son abnégation, et pour son exemple. Viva pur la ~~cioccolata~~ Sonia e colui che l'a inventata !

Joëlle Boretto pour son aide constante, son grand dévouement et ses conseils efficaces et constructifs pour améliorer mes présentations orales. J'espère qu'elle jugera que ses efforts ont porté.

Hélène Dutartre pour sa disponibilité et pour son aide très efficace, et pour avoir vaincu mes préjugés négatifs envers l'immunologie par sa pédagogie et son enthousiasme.

Marie-Pierre Egloff pour son aide constante, très agréable et très appréciée.

Francois Ferron pour son aide répétée, son inspiration, et sa disponibilité sans limites.

Delphine Benarroch pour sa disponibilité, sa gentillesse inspirée et son aide logistique tout au long de cette thèse.

Kenth Johannsson pour nos moments de travail ensemble, qui m'ont beaucoup appris, et pour son aide et son inspiration.

Barbara Selisko pour sa gentillesse et sa disponibilité.

Boulbaba Selmi pour son aide patiente lors de la mise en forme de cette thèse.

Sylvie Liabeuf pour son aide et son dévouement.

Véronique Receveur, Herman Van Tilbeurgh et Nathalie Declerck pour leur conseils avisés et leur aide bibliographique.

Valérie Campanacci pour son aide bienvenue lors de la réalisation des toutes dernières expériences (de dichroïsme circulaire).

Eric Blanc et Philippe Cantau pour leur dévouement

Luciana de Stephanis ainsi que France Chassary pour leur aide constante, efficace, et leur grande disponibilité

Je tiens aussi à remercier les autres membres du laboratoire, passés ou présents, et en particulier Jérôme Deval, Jean-Louis Romette, Magali Sophys, Karine Alvarez et Ling Peng pour leur gentillesse et leur disponibilité.

Bonne chance à tous pour la suite de vos carrières, et en particulier, de vos thèses.

Merci à leurs auteurs pour leurs nombreux conseils très précis, très argumentés, et souvent totalement contradictoires, sur la forme protocolaire que devait prendre cette thèse et sa présentation orale, et en particulier, sur la place et la nature des remerciements.

Certains des nombreux événements que nous avons vécus ensemble et l'ambiance toute particulière du milieu de la recherche seront peut-être un jour immortalisés sous forme de livre, de film, ou ?…

Résumé

Le virus de la rougeole est un virus à ARN négatif, membre de la famille des *Paramyxoviridae*. Le complexe de réplication viral comprend trois protéines principales : la polymérase à ARN ARN-dépendante (L), la nucléoprotéine (N) et la phosphoprotéine (P). Cette thèse présente une étude structurale, par des méthodes biochimiques et biophysiques, de la phosphoprotéine et de la nucléoprotéine produites dans la bactérie *Escherichia coli*. J'ai étudié les déterminants de la polymérisation de N et identifié un variant intéressant pour l'étude de N par cristallographie aux rayons X. J'ai aussi mis au point la purification d'un complexe de N et P, prélude à son étude structurale. Par ailleurs, j'ai montré que la partie N-terminale de P est intrinsèquement désordonnée. J'ai pu étendre ce résultat aux protéines P de nombreux virus apparentés par une étude bio-informatique. J'ai ensuite démontré l'importance du désordre structural au sein du complexe réplicatif des *Paramyxoviridae* (à la fois par son abondance et d'un point de vue fonctionnel). Au-delà de leur intérêt pratique immédiat, ces résultats indiquent que la machinerie réplicative de ces virus pourrait constituer un système modèle de l'étude du désordre structural dans les protéines. De plus, ils soulèvent de nombreuses questions et sont donc un prélude à une refonte de notre vision du complexe réplicatif de ces virus dont l'étude revêt une importance majeure en santé humaine.

Structural study of the proteins of the replicative complex of measles virus

Summary

Measles virus is a negative-strand RNA virus, member of the family *Paramyxoviridae*. Its replicative complex is composed of three main viral proteins : the RNA-dependent RNA polymerase (L), the nucleoprotein (N) and the phosphoprotein (P). This thesis presents a structural study, using biochemical and biophysical methods, of the phosphoprotein and of the nucleoprotein produced in the bacteria *Escherichia coli*. I have studied the regions involved in the polymerisation of N and identified a variant of interest for the study of N using X-ray crystallography. I have also set up the purification of a complex of N and P, a prelude to its structural study. Besides, I have shown that the N-terminal moiety of P is intrinsically disordered. I was able to extend this result to the P proteins of numerous related virus using biocomputing methods. Then, I have demonstrated the importance of structural disorder within the replicative complex of *Paramyxoviridae* (both from a functional point of view and because of its abundance). Beyond their immediate practical interest, these results indicate that the replicative machinery of these viruses might constitute a model system for the study of structural disorder in proteins. Furthermore, they raise numerous questions, and command a rethink of our vision of the replicative complex of these viruses, the study of which has a major importance in human health.

Mots-clés :

virus de la rougeole, *Morbillivirus*, *Paramyxoviridae*, *Rhabdoviridae*, phosphoprotéine, nucléoprotéine, polymérase virale, protéines recombinantes, désordre structural, protéines non structurées, agrégation, polymérisation, mutagenèse dirigée, microscopie électronique.

Discipline : sciences de la vie

Groupe de Virologie & Pathologie Humaine
Laboratoire **d'Architecture et Fonction des Macromolécules Biologiques** (AFMB)
UMR 6098, CNRS et Universités d'Aix-Marseille I et II
31, chemin Joseph Aiguier
13009 Marseille